

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 13:54:35 ; Search time 120 seconds  
(without alignments)  
9489,090 Million cell updates/sec

Title: US-10-003-354-3  
Perfect score: 3713  
Sequence: 1 attaacagcgctgtagg.....aaacttaagattatta 3713

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCrus\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.4	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
2	51.4	1.4	7218	1 US-08-232-463-14	Sequence 14, Appl
3	42.6	1.1	731	1 US-08-451-405A-2	Sequence 2, Appl
4	38.2	1.0	87350	3 US-08-781-891-79	Sequence 79, Appl
5	38.2	1.0	87543	4 US-09-791-211-3	Sequence 3, Appl
6	37.8	1.0	1086	4 US-09-134-001C-1208	Sequence 1208, Ap
7	37.6	1.0	624	4 US-09-397-992A-3	Sequence 3, Appl
8	36	1.0	624	4 US-09-397-992A-6	Sequence 6, Appl
9	36	1.0	1983	4 US-09-453-702B-36	Sequence 36, Appl
10	36	1.0	53526	3 US-08-658-136-2	Sequence 2, Appl
11	36	1.0	53577	3 US-08-658-136-1	Sequence 1, Appl
12	35.8	1.0	549	4 US-09-118-442-26	Sequence 26, Appl
13	35.8	1.0	549	4 US-09-677-064-26	Sequence 26, Appl
14	35.8	1.0	1164	4 US-09-134-001C-2534	Sequence 2534, Ap
15	35.8	1.0	1166	4 US-09-072-596-323	Sequence 323, App
16	35.8	1.0	3095	6 5231168-1	Patent No. 5231168
17	35.8	1.0	19124	2 US-08-487-826B-13	Sequence 13, Appl
18	35.6	1.0	26385	4 US-08-961-527-3	Sequence 3, Appl
19	35.4	1.0	1590	3 US-08-747-221B-23	Sequence 23, Appl
20	35.4	1.0	1590	4 US-09-005-051-23	Sequence 23, Appl
21	35.4	1.0	1650	3 US-08-747-221B-21	Sequence 21, Appl
22	35.4	1.0	1650	3 US-08-747-221B-22	Sequence 21, Appl
23	35.4	1.0	1650	4 US-09-005-051-21	Sequence 22, Appl
24	35.4	1.0	1650	4 US-09-005-051-22	Sequence 22, Appl
25	35.4	1.0	1792	3 US-08-747-221B-18	Sequence 18, Appl
26	35.4	1.0	1792	3 US-08-747-221B-20	Sequence 20, Appl
27	35.4	1.0	1792	4 US-09-005-051-18	Sequence 18, Appl

28	35.4	1.0	1792	4 US-09-005-051-20	Sequence 20, Appl
29	35.4	1.0	2219	5 PCT-US93-07213-4	Sequence 4, Appl
30	35.4	1.0	2870	5 PCT-US93-07213-3	Sequence 3, Appl
31	35.4	1.0	3102	5 PCT-US93-07213-1	Sequence 1, Appl
32	35	0.9	865	4 US-09-040-984-71	Sequence 71, Appl
33	35	0.9	865	4 US-09-123-912-71	Sequence 71, Appl
34	35	0.9	865	4 US-08-356-340-71	Sequence 71, Appl
35	35	0.9	1969	2 US-08-786-555-1	Sequence 1, Appl
36	35	0.9	1969	2 US-08-786-555-1	Sequence 1, Appl
37	35	0.9	2566	2 US-08-488-940-19	Sequence 19, Appl
38	34.6	0.9	35081	2 US-08-752-760A-1	Sequence 1, Appl
39	34.4	0.9	1766	4 US-08-961-527-235	Sequence 235, App
40	34.2	0.9	2659	3 US-08-749-522-1	Sequence 1, Appl
41	34.2	0.9	2755	3 US-08-749-522-2	Sequence 2, Appl
42	34.2	0.9	56516	2 US-08-996-306-1	Sequence 1, Appl
43	34.2	0.9	56516	4 US-09-338-907-1	Sequence 1, Appl
44	34.2	0.9	56516	4 US-09-218-207-1	Sequence 1, Appl
45	34	0.9	421	4 US-08-642-274D-55	Sequence 55, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)856-9300  
TELEFAX: (703)853-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F1s  
US-08-232-463-14  
Query Match 1.5%; Score 55.4; DB 1; Length 7218;





```
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31206
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31592
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 33095
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 33160
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34066
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34072
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 36816
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 39020
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42164
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42459
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46808
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46823
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46826
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 47291
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52786
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52787
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
```

```
NAME/KEY: unsure
LOCATION: 68697
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68718
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68733
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68739
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 69785
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79134
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79198
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 86336
OTHER INFORMATION: unknown
OTHER INFORMATION: unknown
US-09-791-211-3
```

```
Query Match 1.0%; Score 38.2; DB 4; Length 87543;
Best Local Similarity 47.0%; Pred. No. 7.8;
Matches 118; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
```

```
QY 3449 TTTTTCACAGAGAAAGAAATTCCTGCTATTTTTCACATATTTACTATTTATGANG 3508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 45874 TTTATGTATGAGAAATATCATACATAGTTATTTATCTAATTTCTAATACCCAG 45933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3509 TATTTAGCTTTTATTAAGACAGAGTCTGTAGGCGTGGAGGAATATTTGAGGA 3568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 45934 GAAGAAATATATTTATGAAATGATGTCTGTATTTACTGTTTGAGATGCGAAAT 45993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3569 GGGCTGGCTCTTAGGAAAGAAATGGGAGCAACATTTTATTAAGTGTACTATTTGC 3628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 45994 TTAATGTCTTTCTAAMACGACCTGTCAGTAGAATACAAATGAGCTACATATGCA 46053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3629 CTCCTCTGTATTTCTCAAAATGGCAATACATATTAATATGATATGTTTAAAT 3688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46054 TTTTAAATTTCTAGTAGCCACATTTTAAAGTAATGATCAATTTATTTGATTAAT 46113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3689 GTATTAACCTT 3699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46114 ATAATTTAATT 46124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 6
US-09-134-001C-1208
; Sequence 1208, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1208
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1208
```

	Query Match	1.0%;	Score 37.8;	DB 4;	Length 1086;	
	Best Local Similarity	55.8%;	Pred. No.0.51;	Mismatches	72; Conservative	0; Gaps 0;
QY	2916 TTTTCGAAATTTGAATAAATATTCTGGTGTCTTTTTCACAAAGTGGAAGACTTACCAAATGAAT					2975
Db	385 TATTGCATTATCAATATATATGAGATTCTTGTCATGCTTAATCATATTTACAAAATTAAT					444
QY	2976 TTTAGATCATTCTCCAGAGGAGATTTTTTTTGGCTCTTCTCATCTTTGCCAACAGTGTCT					3035
Db	445 TTTAGATTGTCAAATGTAATTAATTTGGCTGCTTTAGACATTATATGTCACAACAACTTAAT					504
QY	3036 CCTGTTGT 3044					
Db	505 CTACATTAT 513					
	RESULT 7					
	US-09-397-992A-3/c					
	: Sequence 3, Application US/09397992A					
	: Patent No. 6329175					
	GENERAL INFORMATION:					
	: APPLICANT: Conklin, Darrell					
	: APPLICANT: Grant, Francis J.					
	: APPLICANT: Rixson, Mark W.					
	: APPLICANT: Kindsvogel, Wayne					
	: TITLE OF INVENTION: Interferon-epsilon					
	: FILE REFERENCE: 98-46					
	: CURRENT APPLICATION NUMBER: US/09/397,992A					
	: CURRENT FILING DATE: 1999-09-16					
	: PRIOR APPLICATION NUMBER: 60/101,012					
	: PRIOR FILING DATE: 1998-09-18					
	: PRIOR APPLICATION NUMBER: 60/118,578					
	: PRIOR FILING DATE: 1999-02-05					
	: PRIOR APPLICATION NUMBER: 60/142,766					
	: PRIOR FILING DATE: 1999-07-08					
	: NUMBER OF SEQ ID NOS: 33					
	: SOFTWARE: FastSeq for Windows Version 3.0					
	: SEQ ID NO 3					
	: LENGTH: 624					
	: TYPE: DNA					
	: ORGANISM: Artificial Sequence					
	: FEATURE:					
	: OTHER INFORMATION: This degenerate sequence encodes the amino acid					
	: NAME/KEY: variation					
	: LOCATION: (1)...(624)					
	: OTHER INFORMATION: n is any nucleotide					
	US-09-397-992A-3					
	Query Match	1.0%;	Score 37.6;	DB 4;	Length 624;	
	Best Local Similarity	27.2%;	Pred. No.0.4;	Mismatches	72; Conservative	60; Indels 3; Gaps 1;
QY	3421 TTTCTGGCTGCGTCTTGCTGCCATGTTTTTTCAGAAGAAGAAATTTCTGCTATT					3480
Db	310 TTTCTGTCACANCCRFCNARNRNSMDARTTNCGCKRAAARNRNSWRADATTYTGTGNARCA					251
QY	3481 TTTTTCATATAATTACTATTATGATGATGATTTAAAGTGTTTTATTAAAGACAGACTTCTG					3540
Db	250 TTTCTGTGARATATNCNARNRGTRTNCCTYYT--GRTAYGYGNGNSMNAHNSWMT					194
QY	3541 TTAGGGGNGGGAGGAATTTGAGGAGAGGCGTGCTTTAGGGAAGAAGATGGGGAAGC					3600
Db	193 TTTGNGNARNRRARRARTTYTNNCKRTGNGNARCAATGYTGDAITNSNANNGTYTGNA					134
QY	3601 AACATTTTATTAACTGTACTATTGCTCTACTCTTGATGTGTCAGAAATGCAATA					3660
Db	133 RTTBTNTMARNRARYTNNARNRSWTGYGTGTTNACTGNCKYIGYGRADAIDATNARYT					74
QY	3661 CAAATATAAAGTATATATGTTT 3685					

```

Db      73  TNAARTCNARNSWRADATNGTNGT 49

RESULT 8
US-09-397-992A-6/c
: Sequence 6, Application US/09397992A
: Patent No. 6329175
:
GENERAL INFORMATION:
: APPLICANT: Konklin, Darrell
: APPLICANT: Grant, Francis J.
: APPLICANT: Rixon, Mark W.
: APPLICANT: Kindsvogel, Wayne
: TITLE OF INVENTION: Interferon-epsilon
: FILE REFERENCE: 98-46
:
CURRENT APPLICATION NUMBER: US/09/397,992A
CURRENT FILING DATE: 1999-09-16
PRIOR APPLICATION NUMBER: 60/101,012
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/118,578
PRIOR FILING DATE: 1999-02-05
PRIOR APPLICATION NUMBER: 60/142,766
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 624
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: This degenerate sequence encodes the amino acid
OTHER INFORMATION: sequence of SEQ ID NO:5.
NAME/KEY: variation
LOCATION: (1)...(624)
OTHER INFORMATION: n is any nucleotide
US-09-397-992A-6

Query Match      1.0%, Score 36; DB 4; Length 624;
Best Local Similarity 26.8%; Pred. No. 1.2;
Matches 71; Conservative 60; Mismatches 131; Indels 3; Gaps 1;

OY 3421 TTTCTGGCGTGGCGTCTGTCGTCATGTTTTCACAGAGGAAGAAATCTGCTAFT 3480
Db 310 TTTCTCCCAACCCRCRCNARNNSMDATRTTNGCCNKAANARNNSWRADATTTGTGNACA 251
OY 3481 TTTTTCATATAATTTACTATTATGATGATGATTTAAAGTGTTTTATTAAGACAGACTCTG 3540
Db 250 TTTCTGTGARNADPTTNCNARNRCRTGNCCTTCTT--GRTAATGTYGNGNSNARNNSWT 194
OY 3541 TTAGGGGGGGGGAATATTTAGGAGGAGGCGTGGGCTTGAGGAAAGAAATGGGGAGC 3600
Db 133 TTTGGGNARNRRNRRARATTTTNCRTGNGNARNCATGTGTGDTNSNARNNGTGTGNA 134
OY 3601 AACATTTTATTAAGTGTACTATTATTTGCTCTACTTTGTATTTGTCCAGAAATGCCAATA 3660
Db 133 RYTTTNTNARNRRYTTNARNNSWTCTYGTHTTACCTGKNCKYTYGTAADATDATNARYT 74
OY 3661 CATATATAAGCATATATATGCTTT 3685
Db 73 TNAARTCNARNSWRADATNGTNGT 49

RESULT 9
US-09-453-702B-36
: Sequence 36, Application US/09453702B
: Patent No. 6365723
:
GENERAL INFORMATION:
: APPLICANT: Blattner, Frederick R.
: APPLICANT: Burland, Valerie
: APPLICANT: Berna, Nicole T.
: APPLICANT: Plunkett, Guy
: APPLICANT: Welch, Rod
:
TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157
NUMBER OF SEQUENCES: 265

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.44mb storage

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION NUMBER:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166

SEQUENCE CHARACTERISTICS:  
LENGTH: 1983  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 36:  
US-09-453-702B-36

Query Match           1.0% ; Score 36; DB 4; Length 1983;  
Best Local Similarity 52.7%; Pred. No. 2.7; Gaps 0,  
Matches 78; Conservative 0; Mismatches 70; Indels 0,

OY 3395 TGTCTATATTAGACACCACCAGTTTCGTCGCCGTCGTCGGCATGTTTTT 3454  
||| ||| ||| | | ||| ||| ||| ||| ||  
Db 477 TGTAATAATTTAAAAAATCAATTACTCTTCCGCCTAATTTTTCTTGAGAATTCATT 536

OY 3455 ACAAGAAGAAAGATTTCTGCTATTTTTTTTTCATAATTACTATTNATGATTTTA 3514  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||  
Db 537 GGCGCATGAATATTTCTTTTAAAGCGATGATGTGAAAATATCAAATATTCGATACCTGT 596

OY 3515 AGTGTTTTAAAGACAGAGTTCTGT 3542  
| ||| ||| ||| ||| ||| ||  
Db 597 TTtAtTTTAATAAcTgTTgAtTTTcTt 624

RESULT 10  
US-08-658-136-2/c  
Sequence 2, Application US/08658136  
Patent No. 6071717

GENERAL INFORMATION:  
APPLICANT: KLINGER, KATHERINE W  
APPLICANT: LINDS, GREGORY M  
APPLICANT: BURN, TIMOTHY C  
APPLICANT: CONNORS, TIMOTHY D  
APPLICANT: DACIKOWSKI, WILLIAM  
APPLICANT: GERMINO, GREGORY  
APPLICANT: OLAN, FENG  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: ONE MOUNTAIN ROAD  
CITY: FRAMINGHAM  
STATE: MASSACHUSETTS  
COUNTRY: USA

```

? ZIP: 01701
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/658,136
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: LASSEN, ELIZABETH
? REGISTRATION NUMBER: 31,845
? REFERENCE/DOCKET NUMBER: GEN4-17.8
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 508-872-8400
? TELEFAX: 508-872-5415
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35326 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match          1.0%; Score 36; DB 3; Length 53526;
Best Local Similarity 51.2%; Fred.No. 26;
Matches   84; Conservative 0; Mismatches 80; Indels 0; Gaps 0.

QY    241 AGCAGCCGGCTCGACGTGTCGTGAGGAGAGGCCCGAGAGGGCCGGGAGTGCCTCCACAGA 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB    35344 AGCAGGGGGATGGATGGAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 35285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY    301 CGCGGGTTCTGTAAGAAGACGTTGGAGATTTCGATTCGAGAGAGAAACCAGATT 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB    35284 GCGGCGCATGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 35225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY    361 GAAGAAGAGCCAGCGCGCTGAGGGGAGGGGGGCTGCTAAGATGG 404
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB    35224 GGAGGAGAGGAGGGGGGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 35181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-08-658-136-1/c
? Sequence 1, Application US/08658136
? Patent No. 6071717
? GENERAL INFORMATION:
? APPLICANT: KLINGER, KATHERINE W
? APPLICANT: LANDES, GREGORY M
? APPLICANT: BURN, TIMOTHY C
? APPLICANT: CONNORS, TIMOTHY D
? APPLICANT: DACROWSKI, WILLIAM
? APPLICANT: GERMINO, GREGORY
? APPLICANT: OTAN, FENG
? TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
? NUMBER OF SEQUENCES: 58
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: GENZYME CORPORATION
? STREET: ONE MOUNTAIN ROAD
? CITY: FRAMINGHAM
? STATE: MASSACHUSETTS
? COUNTRY: USA
? ZIP: 01701
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/658,136
? FILING DATE:
? CLASSIFICATION: 435

```



PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2534  
LENGTH: 1164  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2534

Query Match 1.0%; Score 35.8; DB 4; Length 1164;  
Best Local Similarity 49.7%; Pred. No. 2.1;  
Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 957 AAGAGCGGAAATTTCTGCAAGATGCTTCCAGAGATACATGACCAACCAAGAC 1016  
DB 759 AACATCAGTAATCTGACCAATCATTTTCTTCAACATTTCTTCTACTGTGACTGTC 700  
QY 1017 CTCGACTTTGCTCCCTAAATTCATGACTGTACTGTGACAGGAGGTGCAAGACA 1076  
DB 699 TTTCATTTAAAGGTGTTAAGTCTGCCAAGCTTCTGAGAGTAAGTTGAGGTTTAATAC 640  
QY 1077 TTCGATTTGGTGTATACATCTTTTACCAAGATCGGTAATAATCATATCAATATG 1136  
DB 639 TTCATCTGTATTAATAATTTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 580  
QY 1137 ACC 1139  
DB 579 ACC 577

RESULT 15  
US-09-072-596-323/C  
Sequence 323, Application US/09072596  
Patent No. 6458366

## GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonia  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESS: SPED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,596  
FILING DATE: 05-MAY-1998

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 323:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1166 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-09-072-596-323

Query Match 1.0%; Score 35.8; DB 4; Length 1166;  
Best Local Similarity 29.1%; Pred. No. 2.1;  
Matches 83; Conservative 54; Mismatches 148; Indels 0; Gaps 0;

QY 163 GCGGTTAACTGTGAGAGGGGTGCGGAGACGATTTCTTCCCATGGCAGGCAATGT 222  
DB 559 GAGGGGAATNMGTSGKRGAGADGMTSGKRAATGCTGCMAGNKTGAGRTGNSGG 500  
QY 223 GTGGCTTGAGCTGTGTCAGAGCCGCTCGACTGTCTGAGGAGGCGCCGAGGGGCG 282  
DB 499 GKKKAAAGWACCGGANKKAFRTGRKXMGAGGGRCAFGTGKMTGAGRSCSG 440  
QY 283 GCGAGGTGCCCAACAGAACCGCGTTCTGTAAGAGACCTTGGAAGATTCATTCGAG 342  
DB 439 KKGGRATGACGAGGGRRTGWSMGRTWCGGGGGKNNRAKTASTCCNNGRAKAGASACRKK 380  
QY 343 AAGAGGAAGAACCGGATTTGAAGAGAGCCGCTGAGGGGAGAGGGGTGCTAAGAT 402  
DB 379 SAKTTSMGKSSAGNKKAAAGAMCNAARWNNCCGTTSCCGRGTRRRSAAACCSACAT 320  
QY 403 GCGGTGCGCTCCTCCGGCCGTCGTTGCTGCTGCTTTTCATC 447  
DB 319 CCCGTTACCGGTATTTGAACCAACCCGCGGTGAGRTTATYACC 275

Search completed: January 8, 2003, 18:17:00  
Job time : 632 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 12:53:14 ; Search time 716 Seconds  
(without alignments)  
11678.310 Million cell updates/sec

Title: US-10-003-354-3  
3713  
Sequence: 1 attacagcgccgtgttagg.....aaacttaatgattattta 3713

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3702	99.7	3714	23	AA573986
2	2014	54.2	2133	23	AA573982
3	1852	49.9	2010	23	AA573983
4	1325.2	35.7	1992	23	AA573985
5	581.6	15.7	2764	18	AA843358
6	469.8	12.7	598	22	AA62045
7	469.8	12.7	598	22	AAK10359
8	469.8	12.7	598	22	AAK36262
9	469.8	12.7	598	22	AA141984

10	408	11.0	1291	22	AAH33511
11	408	11.0	1899	24	ABL65517
12	357.4	9.6	1795	23	ABL06531
13	346.6	9.3	3309	23	ABL15963
14	346.6	9.3	6699	23	ABL15962
15	324.8	8.7	674	22	AAH73319
16	324.8	8.7	674	22	AAH73320
17	303.6	8.2	399	21	AAAC04304
18	285	7.7	346	22	ABA74539
19	285	7.7	346	22	AAK23006
20	285	7.7	346	22	AAK49183
21	285	7.7	346	22	AA155018
22	250.6	6.7	885	23	AA568858
23	179	4.8	4687	23	ABL06530
24	164.8	4.4	300	18	AA84371
25	159	4.3	449	23	AA568856
26	159	4.3	449	23	AA573984
27	101.6	2.7	564	22	AAH34797
28	98.6	2.7	153	18	AA84370
29	91.6	2.5	3743	24	ABL69117
30	85.8	2.3	1813	21	AAAC39125
31	79.8	2.1	1099	21	AAAC41576
32	79.8	2.1	1707	21	AAAC41671
33	77	2.1	118	18	AA84369
34	75	2.0	2481	21	AAAC48385
35	75	2.0	2580	21	AAAC42792
36	72.2	1.9	3176	22	AAH42605
37	72	1.9	1219	23	ABL18985
38	70.2	1.9	131	18	AA84368
39	65.8	1.8	174	24	ABN26423
40	63.4	1.7	2872	22	AAF93889
41	61.4	1.7	211	18	AA84372
42	60	1.6	60	24	ABN41013
43	58	1.6	183	21	AAAC09007
44	50	1.3	537	23	ABV45606
45	50	1.3	2095	23	ABV30111

#### ALIGNMENTS

RESULT 1	AA573986	standard; cDNA; 3714 BP.
ID	AA573986	
AC	AA573986;	
DT	13-FEB-2002	(first entry)
DE	DNA encoding novel human diagnostic protein #9790.	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.	
OS	Homo sapiens.	
PN	WO200175067-A2.	
XX	11-OCT-2001.	
PD	30-MAR-2001; 2001WO-US08631.	
PF	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX	(HYSE-) HYSEQ INC.	
PA	Dmanac RT, Liu C, Tang YT;	
PI	WPI: 2001-639362/73.	
DR	P-PSDB; ABG09799.	
XX	New isolated polynucleotide and encoded polypeptides, useful in	





XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
XX	
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HXSE-) HXSEQ INC.
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI, 2001-633962/73.
XX	
XX	P-PsDB; ABG09795.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
PS	Claim 1; SEQ ID No 9786; 103bp; English.
XX	
XX	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridization probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AA564197-AA594564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 2133 BP; 524 A; 524 C; 559 G; 526 T; 0 other;
XX	
XX	Query Match 54.2%; Score 2014; DB 23; Length 2133;
XX	Best Local Similarity 98.2%; Pred. No. 0;
XX	Matches 2073; Conservative 0; Mismatches 0; Indels 39; Gaps 2;
QY	1 ATTACAGCGCGTGTAGGAAGACGAGGAGGAGCGTTCCTCTTGGACATTTTCA 60
DB	
DB	22 ATTACAGCGCGTGTAGGAAGACGAGGAGGCGTTCCTCTTGGACATTTTCA 81
QY	61 TGCCTCGTTTTTTTCAATGATGCGTTGCTGGCGCGCAAGCTCCACAGCACTTA 120
DB	
DB	82 TGCCTCGTTTTTTTCAATGATGCGTTGCTGGCGCGCAAGCTCCACAGCACTTA 141
QY	121 ACCTTACTTCTGTGTAAGGGGAAAGTAATGCCCTGTGAAAGCGGTAACTTGGAG 180
DB	
DB	142 AGCTTACTTCTGTGTAAGGGGAAAGTAATGCCCTGTGAAAGCGGTAACTTGGAG 201
QY	181 GGGGTCGSGAGCGTGAATCTTCCCATGCCAAGCGAATGCTGTGGCTTGAAGCTGTCC 240
DB	
DB	202 GGGGTCGSGAGCGTGAATCTTCCCATGCCAAGCGAATGCTGTGGCTTGAAGCTGTCC 261

QY	241	AGAGCCGCGCTCGACGCTGCTGAGGGAGAGCCCGGAGAGGGGCGGGAGAGTGGCCACAGAA	300
Db	262	AGGAGCCGGGCTCGACGCTGTGAGGGAGGCGCGGAGGGGGGGAGGGTGGCCACAGAA	321
QY	301	CGCGGGTCTGTAAAGAGAGCTTGGGAGATTGATTCGAAAGAGAGAGAGACGGATTT	360
Db	322	CGCGGGTCTGTAAAGAGAGCTTGGGAGATTGATTCGAAAGAGAGAGAGACGGATTT	381
QY	361	GAAGAAGAGCCAGGCGCCGTAGAGGGGAGGGGGGTGTAAAGATGGGTGGGCTCTCCGG	420
Db	382	GAAGAAGAGCCAGGCGCCGTAGAGGGGAGGGGGGTGTAAAGATGGGTGGGCTCTCCGG	441
QY	421	GCCGCGCTTCGGGTGGGTTTTTCATCCATTGATTCGCCGGGCTCCCTCTGACCTTGTG	480
Db	442	GCCGCGCTTCGGGTGGGTTTTTCATCCATTGATTCGCCGGGCTCCCTCTGACCTTGTG	501
QY	481	CT--CAGCATCTGGAAATCAAGAGACCATGCAATGTA-----	516
Db	502	CTCAGCACATCTGGAAATCAAGAGACCATGCAATGCAATGTA-----	561
QY	517	-----GGTGGCTTAATGCTGTGGCAATGCCATCAAGAAATAATGGCCATAG	561
Db	562	TTCTTACATCTCAATGGATGGCTTAATGGCTGTGGCAATGCCATCAAGAAATAATGGCCATAG	621
QY	562	AAGTGTGATTCCTCAGAGAGAGACAATATAAAGAGACACCTCATCAGCCTTGAAGAG	621
Db	622	AAGTGTGATTCCTCAGAGAGAGACAATATAAAGAGACACCTCATCAGCCTTGAAGAG	681
QY	622	TGCCATCCAGTTAGGCATTACCAACACTGTGGGAGCCTGATAGCCAAACAGAGCTGA	681
Db	682	TGCCATCCAGTTAGGCATTACCAACACTGTGGGAGCCTGATAGCCAAACAGAGCTGA	741
QY	682	TGTCTCTATGCAAAATTTCTACGAGGTGAGATATCTTCTTCCAGTGAAGGGAGCA	741
Db	742	TGTCTCTATGCAAAATTTCTACGAGGTGAGATATCTTCTTCCAGTGAAGGGAGCA	801
QY	742	CCTGACCCCTGCTCATCAGTCAATGATGACTTTGTTCAAGACCTATGCACTGTGGCTT	801
Db	802	CCTGACCCCTGCTCATCAGTCAATGATGACTTTGTTCAAGACCTATGCACTGTGGCTT	861
QY	802	CCGACTACTCCGGAGAGTATTGGTATCCGGCCGAGATGATTAATCTGTATTCCTCTGAG	861
Db	862	CCGACTACTCCGGAGAGTATTGGTATCCGGCCGAGATGATTAATCTGTATTCCTCTGAG	921
QY	862	TGAGCCGCTGATTAAGCTGTGAGCTGTGAGGTAGTGTTCCATTCTATGTGTGACAG	921
Db	922	TGAGCCGCTGATTAAGCTGTGAGCTGTGAGGTAGTGTTCCATTCTATGTGTGACAG	981
QY	922	CGAGCATAGTTCATTTATTAAGACAGTCAACATAAGAGGGGGAAATTCGACAAACCT	981
Db	982	CGAGCATAGTTCATTTATTAAGACAGTCAACATAAGAGGGGGAAATTCGACAAACCT	1041
QY	982	GCTTTCAGAGATCAATGAAACCTCAACACGAAGCCCTCGAGACTTGTGCTCAAAATTTGA	1041
Db	1042	GCTTTCAGAGATCAATGAAACCTCAACACGAAGCCCTCGAGACTTGTGCTCAAAATTTGA	1101
QY	1042	TGGACTGTACTGTGTGCAAGGAGGTGGCAAGAACATTCGGAATGTGGTGATATCAAACT	1161
Db	1102	TGGACTGTACTGTGTGCAAGGAGGTGGCAAGAACATTCGGAATGTGGTGATATCAAACT	1161
QY	1102	TTTACCAAGATCCGTAATAAATGCAATATCAAAATTTGACTTCMAAGGCTCAACCTACAAAG	1161
Db	1162	TTTACCAAGATCCGTAATAAATGCAATATCAAAATTTGACTTCMAAGGCTCAACCTACAAAG	1221
QY	1162	GGGGGCTTCCAGAAAGAGCAGAGAAAGCTTCCACATTTAAAGACCTAGACTTCTT	1221
Db	1222	GGGGGCTTCCAGAAAGAGCAGAGAAAGCTTCCACATTTAAAGACCTAGACTTCTT	1281
QY	1222	ACAACAATCCCTATAGTGTCTTTTGTGATGCTGACATGTACAAAGGCTCTGTGTAAGAC	1281
Db	1282	ACAACAATCCCTATAGTGTCTTTTGTGATGCTGACATGTACAAAGGCTCTGTGTAAGAC	1341
QY	1282	CCTGAGGCTACTGTTTGTGTGCTGCAAGACTTCAAGATTAATGAGATTACAGGCTTGTAT	1341



OY	361	GAAGAGAGGCGACGGCGGTGAGAGGGGGGGGGGGGGGGCTCTAAAGTGGGCGGCGGCGTCCGCCG	420
Db	362	GAAGAGAGGCGACGGCGGTGAGAGGGGGGGGGGGGGGGCTCTAAAGTGGGCGGCGGCGTCCGCCG	421
OY	421	GCGCGCTCTCGGGTGGTTTTTTCATCTCTTGTATCCCGGGGTCCCTTCGTACCTTGTG	480
Db	422	GCGCGCTCTCGGGTGGTTTTTTCATCTCTTGTATCCCGGGGTCCCTTCGTACCTTGTG	481
OY	481	CTCAGCATCTGGGAATCAAGAGACCATGGCATCTAGGTGGCTTATGCTCTGTGGCATGCC	540
Db	482	CTCAGCATCTGGGAATCAAGAGACCATGGCATCTAGGTGGCTTATGCTCTGTGGCATGCC	541
OY	541	CATCAATAAATAATAGCCCATAGAAAGTGTGATTCCTCAGAGAGACCAATATATAAAGAC	600
Db	542	CATCAATAAATAATAGCCCATAGAAAGTGTGATTCCTCAGAGAGACCAATATATAAAGAC	601
OY	601	AACCTCATACGCTTGAAAGGTGCCATCCAGTTAGGCAATTCCACACACTGTGGGAGCCT	660
Db	602	AACCTCATACGCTTGAAAGGTGCCATCCAGTTAGGCAATTCCACACACTGTGGGAGCCT	661
OY	661	GAGTACAAACACAGAGCGTGATGCTCTCATGTGCAAAATTTTCACGTGTGAGATATCTT	720
Db	662	GAGTACAAACACAGAGCGTGATGCTCTCATGTGCAAAATTTTCACGTGTGAGATATCTT	721
OY	721	CTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCTACATCAATGACTTTGTTCAA	780
Db	722	CTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCTACATCAATGACTTTGTTCAA	781
OY	781	GACCTATGCAACCGTGTGGCCTTCCGCTACTTTCGGGAGCGCATTTGGTATCCGGCCGATGA	840
Db	782	GACCTATGCAACCGTGTGGCCTTCCGCTACTTTCGGGAGCGCATTTGGTATCCGGCCGATGA	841
OY	841	TTACTTGTATTCCTCTGTCAGTGTAGCGCGGTGATTAACCTCTGTAGCTCTGGAGCTAGTG	900
Db	842	TTACTTGTATTCCTCTGTCAGTGTAGCGCGGTGATTAACCTCTGTAGCTCTGGAGCTAGTG	901
OY	901	TTCCCTATCTATGTGTCCAGCGAGATAGTTCATTTTAAGCAAGTCCAAACTTAAGA	960
Db	902	TTCCCTATCTATGTGTGTCCAGCGAGATAGTTCATTTTAAGCAAGTCCAAACTTAAGA	961
OY	961	GGCGGAATTTCTGAGAGAGCTGCTTCACAGATACATCAATGAACCTCAACCGAACCCTCG	1020
Db	962	GGCGGAATTTCTGAGAGAGCTGCTTCACAGATACATCAATGAACCTCAACCGAACCCTCG	1021
OY	1021	GACCTTGTGCTGCTAAATTTCTATGGAATGTATGTGTGTGCAAGGAGGTGGCAAGAACATTGC	1080
Db	1022	GACCTTGTGCTGCTAAATTTCTATGGAATGTATGTGTGTGCAAGGAGGTGGCAAGAACATTGC	1081
OY	1081	GATTGTGGTGTGTAACAATTTTATTACAGATTCGGTAAATAATGCAATCAATATATGACT	1140
Db	1082	GATTGTGGTGTGTAACAATTTTATTACAGATTCGGTAAATAATGCAATCAATATATGACT	1141
OY	1141	CAAAAGCTCAACTCAACAACGGCGGGCTTCCAGAAAGGCGAGAAAGCCTTTCCAC	1200
Db	1142	CAAAAGCTCAACTCAACAACGGCGGGCTTCCAGAAAGGCGAGAAAGCCTTTCCAC	1201
OY	1201	ATTTAAAGACCTTAACCTTTTACAAAGCAATCCCTGATGCTCTTTTGTGATGCTGCAT	1260
Db	1202	ATTTAAAGACCTTAACCTTTTACAAAGCAATCCCTGATGCTCTTTTGTGATGCTGCAT	1261
OY	1261	GTACAAAGCTCTGTGAAGACCTGCAAGCGTGAAGTGTGTGTGTGCAAGACTTCAGAT	1320
Db	1262	GTACAAAGCTCTGTGTGAAGACCTGCAAGCGTGAAGTGTGTGTGTGCAAGACTTCAGAT	1321
OY	1321	AATGATTTACAGCCTCTTGATGTGTATCCATTAATAGTATCATGTGACAAGCAAGACCCCTT	1380
Db	1322	AATGATTTACAGCCTCTTGATGTGTATCCATTAATAGTATCATGTGACAAGCAAGACCCCTT	1381
OY	1381	AAGCGTGAACACAGTATCACTAGTGTACTGCAAAACCGGCCCCCAAAAGGCTCTGTA	1440
Db	1382	AAGCGTGAACACAGTATCACTAGTGTACTGCAAAACCGGCCCCCAAAAGGCTCTGTA	1441
OY	1441	TTCCAGACCCATGATTCATCCAGGAGAGGCTGACGGGGTGGTACCATGAGACTGA	1500

Db	1442	TTCCACAGCCATGGAATTCATCCAGGAGAGAGCTCGACGGGTGGTACCATGAGACCTGA	1501
Oy	1501	TGACCATATGGGTGGGCATCCCTGCGCGGAATAGTAAGGGGAAGGCTTTCCTTAAT	1560
Db	1502	TGACCATATGGGTGGGCATCCCTCGCGGAATAGTAAGGGGAAGGCTTTCCTTAAT	1561
Oy	1561	TGGCATCATGCACTTCCTACAGTCTTACAGAGTTGTTAAGAAAGTTGGAGCACTCTTGGAA	1620
Db	1562	TGGCATCATGCACTTCCTACAGTCTTACAGAGTTGTTAAGAAAGTTGGAGCACTCTTGGAA	1621
Oy	1621	AGCCCTGGTACATGACGAGAACACTGTCTCATGTGATCGCCAGGCTTCTACGTGAACG	1680
Db	1622	AGCCCTGGTACATGACGAGAACACTGTCTCATGTGATCGCCAGGCTTCTACGTGAACG	1681
Oy	1681	GTTCACAGGCTTCATGTGCAACACAGATTTAAGATTTCCCTTGAAGCCTTTCCTTC	1740
Db	1682	GTTCACAGGCTTCATGTGCAACACAGATTTAAGAAATTTCCCT	1723
Oy	1741	CAAAAAGTTTCGTGCTGCTCATCTTCTCTCGGCGAGCAGGCTCCAGTGCMACTCCTG	1800
Db	1726	-----	1723
Oy	1801	CATTACTTACCAAGCCATCGGTCTCTGGGGAAACAAGCAAGTACACAAGCAAGGCAGA	1866
Db	1726	-----	1723
Oy	1861	AGTGAGCCAGGCGTTCACCTTGGTCGCTCGATGTTTACCTAGACACTCACCTTTGGA	1920
Db	1726	-----GCGTTACCGTTGGTGTGCTGTGATGTTTACCTAGACACTCACCTTTGGA	1774
Oy	1921	GGAATTCAGTGAAGGCTCGCTATTCCTGACCCCAAGTTTCTACCTCTAGTTGGAGAGAC	1980
Db	1775	GGAATTCAGTGAAGGCTCGCTATTCCTGACCCCAAGTTTCTACCTCTAGTTGGAGAGAC	1834
Oy	1981	TTTGCAATTCGTACTACAGTACAACTTGGAAAAGCTGAAGTGCAGAGTCAAGATT	2040
Db	1835	TTTGCAATTCGTACTACAGTACAACTTGGAAAAGCTGAAGTGCAGAGTCAAGATT	1894
Oy	2041	CACCCATTAAAGCGCAAAAGCCTCAGAGACCTGGAACAAGATTCTGCCATCTCTGTATCC	2100
Db	1895	CACCCATTAAAGCGCAAAAGCCTCAGAGACCTGGAACAAGATTCTGCCATCTCTGTATCC	1954
Oy	2101	CAAGATGTCAAGCCCTTGCCCGCAGCAATGCTGAATTTTCTCTACTGATGATCAAA	2156
Db	1955	CAAGATGTCAAGCCCTTGCCCGCAGCAATGCTGAATTTTCTCTACTGATGATCAAA	2010
RESULT 4			
AA573985			
ID	AA573985	standard; cdna; 1992 bp.	
XX	AA573985;		
AC			
XX	13-FEB-2002	(first entry)	
DE			
XX	DNA encoding novel human diagnostic protein #9789.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
XX	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HTSEQ INC.		

XX Drmanac RT, Liu C, Tang YF;  
 XX MPI; 2001-639362/73.  
 DR P-PSDB; ABG09798.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 1: SEQ ID No 9789; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 1992 BP; 496 A; 520 C; 452 G; 524 T; 0 other:  
 Query Match 35.7%; Score 1325.2; DB 23; Length 1992;  
 Best Local Similarity 92.8%; Pred. No. 0;  
 Matches 1427; Conservative 0; Mismatches 98; Indels 13; Gaps 3;

OY 881 TGTAGCTGTGAGCTAGTGGTTCCTTATTCATGTGTCCAGGAGCATGATTCATTAAT 940  
 DB 481 TGTAGCTGTGAGCTAGTGGTTCCTTATTCATGTGTCCAGGAGCATGATTCATTAAT 540  
 OY 941 AAGACAGTCCAACTAAGAGGCGGAATTTCTGCAGAAAGCTCTCCAGGATACATG 1000  
 DB 541 AAGACAGTCCAACTAAGAGGCGGAATTTCTGCAGAAAGCTCTCCAGGATACATG 600  
 OY 1001 AACCTCAACCAAGCCCTGAGCTTGGCTTAAATTCATGAGTGTGAGTGGCAG 1060  
 DB 601 AACCTCAACCAAGCCCTGAGCTTGGCTTAAATTCATGAGTGTGAGTGGCAG 660  
 OY 1061 GCAGGTGGCAAGACATTCGATGTGTGATGAACAATCTTTACAGATCGGTAA 1120  
 DB 661 ACAGGTGGCAAGACATTCGATGTGTGATGAACAATCTTTACAGATCGGTAA 720  
 OY 1121 ATGCATATCAATATGACCTCAAGGCTCAACCTCAAGGCGGCTTCCAGAAAGAG 1180  
 DB 721 ATGCATATCAATATGACCTCAAGGCTCAACCTCAAGGCGGCTTCCAGAAAGAG 780  
 OY 1181 CGAGGAAGCCCTTCCCATATTTAAAGACCTGATTTCAAGACATCCCTGATG 1240  
 DB 781 CGAGGAAGCCCTTCCCATATTTAAAGATCTGATTTCAAGACATCCCTGATG 840  
 OY 1241 CTTTATTTGATGCTGACATGTACAAAGCTCTGTGAAGACCTGACGCTGCTTTG 1300  
 DB 841 CTTTATTTGATGCTGACATGTACAAAGCTCTGTGAAGACCTGACGCTGCTTTG 900  
 OY 1301 GTGCTGCAGAGCTTCAAGATATGATGATGACCTCTGATGATCAATCATATATGAT 1360  
 DB 901 GTGCTGCAGAGCTTCAAGATATGATGATGACCTCTGATGATCAATCATATATGAT 960  
 OY 1361 CATGCAACAAGAGCCCTTAAGACAGTGAAC---ACATGATGATGATGATGATG 1417  
 DB 961 CATGCAACAAGAGCCCTTAAGACAGTGAAC---ACATGATGATGATGATGATG 1020  
 OY 1418 CCGGCCCCCAAAAGGCTCTGTATTCACAGGATGGAATCCAGGAGAGGCTGCA 1477  
 DB 1021 CCGGCCCCCAAAAGGCTCTGTATTCACAGGATGGAATCCAGGAGAGGCTGCG 1080  
 OY 1478 CCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1537  
 DB 1081 CCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 OY 1538 GGGGAAAGGCTCTGATATATGATGATGATGATGATGATGATGATGATGATG 1597  
 DB 1141 GGGGAAAGGCTCTGATATATGATGATGATGATGATGATGATGATGATGATG 1200  
 OY 1598 AAGAGTTGAGACACTTGTGAAAGCCCTGTATGATGAGAGACATGTTCTAGTGCAT 1657  
 DB 1201 AAGAGTTGAGACACTTGTGAAAGCCCTGTATGATGAGAGACATGTTCTAGTGCAT 1260  
 OY 1658 CGCCAGGCTTGTACGCTGAGACGCTTCCAGGCTTCATGTGCAACACAGTATTAAGAG 1717  
 DB 1261 CGCCAGGCTTGTACGCTGAGACGCTTCCAGGCTTCATGTGCAACAGTATTAAGAG 1320  
 OY 1718 ATTCCCTTGAAGCCTTCTCTTCCAAAGGTTTGGCTGCTGCAATCTTCTGCGGCA 1777  
 DB 1321 ATTCCCTTGAAGCCTTCTCTTCCAAAGGTTTGGCTGCTGCAATCTTCTGCGCA 1380  
 OY 1778 GCAGGCTCCAGTGGCAACTCTGATTAATTAACAGCCATGCTCTGAGGAGCAAG 1837  
 DB 1381 ACAGGCTCCAGTGGCAACTCTGATTAATTAACAGCCATGCTCTGAGGAGCAAG 1440  
 OY 1838 GCACAAGTCAACAAGGCAAGTGGAGCCAGGCTTACCTTGTGCTGATGATG 1897  
 DB 1441 TCACAAGT---ATAAGGTGCAAGTGGAGCCAGGCTTACCTTGTGCTGATGATG 1497  
 OY 1898 TTACCTGACAGC-----TCACCTTTGAGGAAATCA 1928  
 DB 1498 TTACCTGACAGC-----TCACCTTTGAGGAAATCA 1535



```

RESULT 5
AAT84358
ID AAT84358 standard; DNA; 2764 BP.
XX
AC AAT84358;
XX
DT 12-NOV-1997 (first entry)
XX
DE STM7.I gene associated with Friedreich's ataxia.
XX
KW STM7.I gene; Friedreich's ataxia; FRDA; neurodegeneration;
KM phosphatidylinositol-4-phosphate 5-kinase; pldinsp5k;
KW transgenic animal; animal model; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 420..2042
FT /tag= a
FT /tag= b
FT polyA_signal 2749..2753
FT /tag= b
XX
PN MO9705234-A2.
XX
PD 13-FEB-1997.
XX
PF 24-JUL-1996; 96WO-GB01786.
XX
PR 28-JUL-1995; 95GB-0015508.
PR 26-JUL-1995; 95GB-0015309.
XX
PA (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
PI Chamberlain S, Doudney CME, Garcia-Valdecasas JJC;
PI Hillermann R, Pook MA;
XX
DR WPI: 1997-145674/13.
XX
DR P-PSDB; AAM00978.
XX
PT New gene, STM7, and new protein causing Friedreich's Ataxia -
PT encodes phosphatidylinositol phosphate kinase isoform; also splice
PT variants and transgenic animals, for diagnosing pre-disposition to
PT Friedreich's ataxia
XX
PS Claim 1; Fig 7; 82pp; English.
XX
PS
XX
CC This isolated DNA molecule comprises the full-length human STM7.I
CC gene associated with Friedreich's ataxia (FRDA). It comprises
CC exons 1-16 of the STM7 gene (see AAT84364-86) and encodes a
CC phosphatidylinositol-4-phosphate 5-kinase (AAM00978) isoform (P5K).
CC The STM7.I sequence was obtained by construction of chromosome 9
CC region-specific cosmid contigs, exon trapping, infant cerebellum
CC cDNA library screening and RACE. STM7 nucleic acids can be used to
CC produce recombinant p5K in recombinant host cells using a claimed
CC method. Probes derived from exons 1-16, partic. 7-11, of STM7 are
CC used in a claimed diagnostic method for determining an inherited
CC predisposition to FRDA. STM7 nucleic acids can also be used to
CC generate transgenic animal models of FRDA and in somatic cell
CC therapy.
XX
XX
SO Sequence 2764 BP; 850 A; 567 C; 573 G; 774 T; 0 other;

Query Match 15.7%; Score 581.6; DB 18; Length 2764;
Best Local Similarity 69.6%; Pred. No. 3.2e-153;
Matches 820; Conservative 0; Mismatches 349; Indels 9; Gaps 2;

```

```

QY 698 TTCACGCTGGTGAAGATATCTTCTTCCAGTGAAGGAGCAACCTGACCTGCTCAT 757
DB 588 TTTATGTGTGGAAGATGTCTTCCAGGAGGAGGAGCATCTGACCCAGCAT 647
QY 758 CACTACATATCTTGTCTTCAAGACCTATGACACTGTCCCTACTCCGGAG 817
DB 648 CACTACCATGCTTGTGATTTAAGACATACGTCATTAAGATTCGATTTTCAAGAA 707
QY 818 CTATTGTATCCGCGCCGATATCTTATCTTCCCTGACAGTGGCGCTGATTGA 877
DB 708 CTTTGTGTATCAAGCTGTGATTAATCTTATCTTCCATCTGACAGTACCTCTTAATGA 767
QY 878 CTCTGTAGCTGTGAGCTAGCTTCCCTATCTATGTGTCCAGCAGATGATTCATT 937
DB 768 CTGTCTAACCTGAGACGATGATCTTGTGTTGTGACAGTGTATGATTTATTC 827
QY 938 ATTAAGACATCTCCAACTTAAGAGGCGGAAATTTCTGACAGAGCTCTTCCAGATCTAC 997
DB 828 ATCAAAACAGTTCAGCAAAAGAGCTGATTTCTCAGAGCTACTGCCAGGCTATTAC 887
QY 998 ATGAACTCAACCAAGAACCTCGACTTGTGCTTAATTTATGAGTACTGATG 1057
DB 888 ATGAATTTAAACCAAGATTCAGAGACTCTTTGCCAAATTTTACGACTGTATGATG 947
QY 1058 CAGGACAGTGGCAAGAACATTCGATTTGTGATGAACAATCTTTACCAAGATCGTA 1117
DB 948 CAATCAGAGGACATTAATATCAGGATTTGTGATGACAACTTTTGCACAGCTCATG 1007
QY 1118 AAATGATATCAAAATATGACCTCAAAAGGCTCAACCTCAACAGCGGCTTCCAGAAA 1177
DB 1008 AGAATGACTTACATTAATGACTTGAAGGCTCAACGATTAAGAGAGATCCCGTAAA 1067
QY 1178 GACGAGAGAGAGGCTCTCCACATTTAAGACTAGTACTCTTCAAGACATCCCTGAT 1237
DB 1068 GAGAGAGAGAAATTCACACCCACATTTAAGACTTATCTTCCGACAGATGACAGAA 1127
QY 1238 GGTCTTTTGTGATGCTGATGATACACGCTCTGTGAAGACCTGACGCTGATG 1297
DB 1128 GGGTTGTATTTTATATACGAAACATACACGCTTATGAACACTTCAGAGAGACTGC 1187
QY 1298 TTGGTGTGAGAGCTTCAAGATTAATGATTAAGCTCTGTAAGTCAATCCATATATA 1357
DB 1188 CGGTGTGATGAAGCTTCAAGATCATGATTAATGATTCCTGTGGATTCATTTCC 1247
QY 1358 GATCATGACACAGAGAGACCTTAAAGCATGAACACAGTACTCATGATTCGAGAA 1417
DB 1248 GACCATTTCCCTCAAAAGAGAAAGAGAGAGACCCCAAAATGTGCTGTAAACGG 1307
QY 1418 CCGGCCCCCAAAAGGCTGTATTCACAGACCATGGAATCCATCCAG-----GGAGAG 1471
DB 1308 ACTGGGATGCAAGAGGCTCTACTCAACAGCATGGAATCTATCCAGGGTCCAGGGAAA 1367
QY 1472 GCTGACGGGGTGTACCATGAGAGACATGACCATTAATGGGTGGCATCCCTGCCCGGAAT 1531
DB 1368 TGTGGAGATGGATTAATCAAGAGAACCCAGACCAATGAGGAGCATTCAGATAAAAGC 1427
QY 1532 AGTAAAGGGGAAAGGCTTCTGCTTTATATGAGCATATACATCTTCAAGTCTTCAAGG 1591
DB 1428 CATAGGGAGAAAGAACTTATTTATTTATGGCATTAATGATCTGCAATCATATAGG 1487
QY 1592 TTTGTTAAGAGTTGGAGACCTTTGGAAAAGCCCTGGTACATGACGAGACACTGTCTCA 1651
DB 1488 TTATATGAAGAGTTAGAACATTCCTGTGAAAGCTTTGTTATGATGGGACACTGTTTCT 1547
QY 1652 GTGCAATGCCAGGCTTCTACGCTGAAGGTTCCAGGCTTCATGTCACACAGATATT 1711
DB 1548 GTTCATGACCAAGCTTTTATGCAAGAGATTTCTTAAGTTCATGAATTCAGAGATTTC 1607
QY 1712 AAGAAGATTC---CCTGAAGCTTCTCTCCAAAAA 1746
DB 1608 AAGAATAATCAAGCTTTGAAGGCTTCACCGCTTAAGAA 1645

```



Query Match	12.7%	Score 469.8	DB 22	Length 598
Best Local Similarity	90.1%	Pred. No. 3e-122		
Matches 527	Conservative	0	Mismatches 52	Indels 6
				Gaps 2
Sequence 598 BP; 149 A; 156 C; 139 G; 154 T; 0 other;				
1331 ACCCTCTTGATGTCATCCATATATAGATCATGACACAGAGAGCCCTTAAGACAGTGA	1390			
1331 ACCTCTTGATGTCATCCATATATAGATCATGACACAGAGAGCCCTTAAGACAGTGA	1390			
2 ACCCTCTTGATGTCATCCATATATAGATCATGACACAGAGAGCCCTTAAGACAGTGA	61			
1391 AC---ACAGTACTGACGTTGATGATGAGACCGGCCCCCAAAAGGCTGTATTCACA	1447			
1391 ACCTCTTGATGTCATCCATATATAGATCATGACACAGAGAGCCCTTAAGACAGTGA	1390			
62 ACTCTTCAAGTGTCAATCAGACACCTCAAAAGCTGGCTCCCAAAAGGCTGTATTCACA	121			
1448 GCCATGGAATTCATCCAGGAGAGAGGCTCGGCTGGGGGACACCATGAGGCGGATGACAT	1507			
122 GCCATGGAATTCATCCAGGAGAGAGGCTCGGCTGGGGGACACCATGAGGCGGATGACAT	181			
1508 ATGGGTGGCATCCCTCCCGGAGTATGTAAGGGGAAAGGCTTCTGTATATATATGGCATC	1567			
182 ATGGGTGGCATCCCTCCCGGAGTATGTAAGGGGAAAGGCTTCTGTATATATGGCATC	241			
1568 ATTGACATCTTCAAGTCTTACAGGTTTGTAAAGAAAGTGGAGCATCTTTGGAAAGCCCTG	1627			

Db	242	ATTGACATTGTACAGTCTTTACACGTTTCTTAAGAAGTTGGAGCACTTCTGGAAAGCCGCG	301
QY	1628	GTCACATGACGGAGACACTGTCTCAGTGCATGCCAGGCTTTTACGCTGAACGGTTCCAG	1687
Db	302	GTACATGTATGGGGACGCGTGTCTCAGTGCATGCCAGGCTTTTACGCTGAACGGTTCCAG	361
QY	1688	CGCTCATGTGTCACACACAGTATTTTAAGAAGTTTCCCTTGAAGCTTCTCTTCCAAAAG	1747
Db	362	CACCTTCATGTCTCAACGCGCAGTATTTTAAGAAGTATCCCTTGAAGCTTCTCTTCCAAAAG	421
QY	1748	TTTCGGTGTGGCTCATCTTCTCTCGGCGAGCAGGCTCCAGTGCAACTCTCGATTACT	1807
Db	422	TTTCGGTGTGGCTCATCTTCTCTCGGCGAGCAGGCTCCAGTGCAACTCTCGATTACT	481
QY	1808	TACACGCCATCGGTCTCTGGGGAACACAAAGCACAAGTGACAAACAAGCAGAAATGGAG	1867
Db	482	TACACGCCATCGGTCTCTGGGGAACACAAAGCACAAGTGACAAACAAGCAGAAATGGAG	538
QY	1868	CCAGGCGTTTACACTTTGGTCTGCTCTGATGTTTACTTACACTCA	1912
Db	539	CCAGGCGTTTACACTTTGGTCTGCTCTGATGTTTACTTACACTCA	583
RESULT 7			
AAK10359			
ID	AAK10359	standard; DNA; 598 BP.	
XX	AC	AAK10359;	
XX	DT	05-NOV-2001 (first entry)	
DE	XX	Human brain expressed single exon probe SEQ ID NO: 10350.	
XX	XX	Human; brain expressed exon; gene expression analysis; probe:	
KM	XX	microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;	
KM	XX	epilepsy; cancer; ss.	
OS	XX	Homo sapiens.	
XX	PN	WO200157275-A2.	
XX	PD	09-AUG-2001.	
PF	30-JAN-2001;	2001WO-US00667.	
XX	PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000;	2000US-0207456.	
PR	30-JUN-2000;	2000US-0608408.	
PR	03-AUG-2000;	2000US-0632366.	
PR	21-SEP-2000;	2000US-0234687.	
PR	27-SEP-2000;	2000US-0236359.	
PR	04-OCT-2000;	2000GB-0024263.	
PA	(MOL-)	MOLECULAR DYNAMICS INC.	
PI	Penn SG,	Hanzel DK, Chen W, Rank DR;	
DR	WPI;	2001-483446/52.	
XX	PT	Single exon nucleic acid probes for analyzing gene expression in human	
XX	XX	brains -	
PS	Example 4;	SEQ ID NO: 10350; 650pp + Sequence Listing; English.	
CC	The present invention provides a number of single exon nucleic acid		
CC	probes which are derived from genomic sequences expressed in the human		
CC	brain. They can be used to measure gene expression in brain cell samples		
CC	which may enable the diagnosis and improved treatment of nervous system		
CC	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,		
CC	epilepsy and cancers. The present sequence is one of the probes of the		
XX	invention.		

Sequence 598 BP; 149 A; 156 C; 139 G; 154 T; 0 other;

Query Match 12.7%; Score 469.8; DB 22; Length 598;  
Best Local Similarity 90.1%; Pred. No. 5e-122;  
Matches 527; Conservative 0; Mismatches 52; Indels 6; Gaps 2;

QY 1331 AGCCTTGTGATGTAATCCATATATATGATCATGCACAGAGAGCCCTTAAGCATGAA 1390  
DB 2 AGCCTTGGGCTGCAATCCAAATATAGATCATGCACAGAGAGCCCTTAAGCATGAA 61  
QY 1391 AC--ACAGTACTGATGTAATCTGCAAGACCGGCCCCCAAAAGGCTGTATTCACA 1447  
DB 62 ACTTTCAAGTGTCAATCGACACTCAAAAGACTGGCTCCCAAAAGGCTGTATTCACA 121  
QY 1448 GCCATGGAATCCATCCAGGAGAGGCTCGAGGGGGTGTACCATGAGACTGATGACAT 1507  
DB 122 GCCATGGAATCCATCCAGGAGAGGCTCGAGGGGGTGTACCATGAGACTGATGACAT 181  
QY 1508 ATGGGTGCATCCCTGCCGAATAGTAAAGGGAAAGGCTTGTATATTTGGCATC 1567  
DB 182 ATGGGTGCATCCCTGCCGAATAGTAAAGGGAAAGGCTTGTATATTTGGCATC 241  
QY 1568 ATTGCATATTCATACAGCTTACAGAGTGTGTAAGAGTTGAGACACTGTGGAAGCCCTG 1627  
DB 242 ATTGCATATTCATACAGCTTACAGAGTGTGTAAGAGTTGAGACACTGTGGAAGCCCTG 301  
QY 1628 GTACATAGCAGGAGACTGTCTCAGTCAATCGCCAGGCTTTCAGCTGGAAGGTTCCAG 1687  
DB 302 GTACATAGCAGGAGACTGTCTCAGTCAATCGCCAGGCTTTCAGCTGGAAGGTTCCAG 361  
QY 1688 CGCTTCATGTGCAACAGATATTAAGAAATTCCTTGAAGCCTTCTCTCCAAAAG 1747  
DB 362 CACTTCATGTGCAACAGATATTAAGAAATTCCTTGAAGCCTTCTCTCCAAAAG 421  
QY 1748 TTTCGGTGTGCTCATCTTCTCTCGGCGAGAGCTTCAGTGGCACTCTCGATTACT 1807  
DB 422 TTTCGGTGTGCTCATCTTCTCTCGGCGAGAGCTTCAGTGGCACTCTCGATTACT 481  
QY 1808 TACCAGCCATCGTCTCTGCGGGAACACAGGCGCAAGTGAACAACAAGGCAAGTGGAG 1867  
DB 482 TACCAGCCATCGTCTCTGCGGGAACACAGGCGCAAGTGAACAACAAGTGGAG 538  
QY 1868 CCAAGGCTTCACCTTGTGCTGCTGATTTTACCTCAGACTCCA 1912  
DB 539 CCAAGGCTTCACCTTGTGCTGCTGATTTTACCTCAGACTCCA 583

RESULT 8  
AAK36262  
ID AAK36262 standard; DNA: 598 BP.

AC AAK36262;  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 10819.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukemia; lymphoma; myeloma; ss.  
OS Homo sapiens.  
XX  
PN W0200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488900/53.  
XX  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX  
XX  
PS Example 4; SEQ ID NO: 10819; 658bp + Sequence Listing; English.  
XX  
XX  
CC The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukemia and myeloma. The present sequence is one of  
XX the probes of the invention.

Sequence 598 BP; 149 A; 156 C; 139 G; 154 T; 0 other;

Query Match 12.7%; Score 469.8; DB 22; Length 598;  
Best Local Similarity 90.1%; Pred. No. 5e-122;  
Matches 527; Conservative 0; Mismatches 52; Indels 6; Gaps 2;

QY 1331 AGCCTTGTGATGTAATCCATATATATGATCATGCACAGAGAGCCCTTAAGCATGAA 1390  
DB 2 AGCCTTGGGCTGCAATCCAAATATAGATCATGCACAGAGAGCCCTTAAGCATGAA 61  
QY 1391 AC--ACAGTACTGATGTAATCTGCAAGACCGGCCCCCAAAAGGCTGTATTCACA 1447  
DB 62 ACTTTCAAGTGTCAATCGACACTCAAAAGACTGGCTCCCAAAAGGCTGTATTCACA 121  
QY 1448 GCCATGGAATCCATCCAGGAGAGGCTCGAGGGGGTGTACCATGAGACTGATGACAT 1507  
DB 122 GCCATGGAATCCATCCAGGAGAGGCTCGAGGGGGTGTACCATGAGACTGATGACAT 181  
QY 1508 ATGGGTGCATCCCTGCCGAATAGTAAAGGGAAAGGCTTGTGTAATTTGGCATC 1567  
DB 182 ATGGGTGCATCCCTGCCGAATAGTAAAGGGAAAGGCTTGTGTAATTTGGCATC 241  
QY 1568 ATTGCATATTCATACAGCTTACAGAGTGTGTAAGAGTTGAGACACTCTTGAAGCCCTG 1627  
DB 242 ATTGCATATTCATACAGCTTACAGAGTGTGTAAGAGTTGAGACACTCTTGAAGCCCTG 301  
QY 1628 GTACATAGCAGGAGACTGTCTCAGTCAATCGCCAGGCTTTCAGCTGGAAGGTTCCAG 1687  
DB 302 GTACATAGCAGGAGACTGTCTCAGTCAATCGCCAGGCTTTCAGCTGGAAGGTTCCAG 361  
QY 1688 CGCTTCATGTGCAACAGATATTAAGAAATTCCTTGAAGCCTTCTCTCCAAAAG 1747  
DB 362 CACTTCATGTGCAACAGATATTAAGAAATTCCTTGAAGCCTTCTCTCCAAAAG 421  
QY 1748 TTTCGGTGTGCTCATCTTCTCTCGGCGAGAGCTTCAGTGGCACTCTCGATTACT 1807  
DB 422 TTTCGGTGTGCTCATCTTCTCTCGGCGAGAGCTTCAGTGGCACTCTCGATTACT 481  
QY 1808 TACCAGCCATCGTCTCTGCGGGAACACAGGCGCAAGTGAACAACAAGGCAAGTGGAG 1867  
DB 482 TACCAGCCATCGTCTCTGCGGGAACACAGGCGCAAGTGAACAACAAGTGGAG 538  
QY 1868 CCAAGGCTTCACCTTGTGCTGCTGATTTTACCTCAGACTCCA 1912  
DB 539 CCAAGGCTTCACCTTGTGCTGCTGATTTTACCTCAGACTCCA 583

RESULT 9  
AAI1984  
ID AAI1984 standard; DNA: 598 BP.  
AC AAI1984;



Query Match 11.0%: Score 408; DB 22; Length 1291;  
Best Local Similarity 67.7%: Pred. No. 2,4e-104;  
Matches 605; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

QY 862 TGACCCGCTGATTTGACTCTGAGCTAGTGGTCCCTATCTATGTCGAG 921  
DB 1 TGAACCTTATAGAACTGTTAACCCCTGAGCCAGTGGATCCCTGTTTGTGACAG 60

QY 922 CGAGATGAGTTCATTTAAGACATCCAAATTAAGAGCGGAAATTTGACAGACT 981  
DB 61 TGATGATGAATTTATCAACAAACAGTTCACACAAAGAACTGATGTTCTTCAGAACT 120

QY 982 GCTTCAGAGATGATGAACTCAACCACTGAGCTTGGCTTGAATTTCTA 1041  
DB 121 ACTGCCAGCTATTTACATGAAATTTAAACCAAGATTCAGAGCTTTTCCAAAATTTTA 180

QY 1042 TGGACTGATCTGTGACAGCAGGTGCGAAGATTCGATTTGATGAACTATCT 1101  
DB 181 CGGACTGATTTGATGCAATGAGGAGCATTAATATCGATTTGTTGATGAACT 240

QY 1102 TTTACCAAGATCGGTAATAATGATATCAATATGACCTCAAGGCTCAACCTACAAACG 1161  
DB 241 TTTCCAGAGCTCCAGTAAGAAATGCAATGATGACTTGAAGGCTCAACGTTAAAGCG 300

QY 1162 GCGGGCTTCCAGAAAGCAGAGAAAGCCCTCTCCACATTTAAAGCTTACTCTT 1221  
DB 301 AAGAGCATCCCGTAAGAGAGAGAGAAATCCAAACCCATTTAAGCCTTAATTCCT 360

QY 1222 ACAAGACATCCCTGATGCTTTTGTGATGCTGACATGATCAAGCTCTGTAAGAC 1281  
DB 361 GCAGACATGACAGAGGCTGTTTGTATGATGAGAAACATCAACGCTTATGAAAC 420

QY 1282 CCTCAGCGTACTGTTGTTGCTGACAGCTCAAGATTAATGATACGCTTGTAT 1341  
DB 421 ACTTCAGAGACATCGCGGCTGTAAGAAAGCTCAAGATGATGATATACCTTCTGTT 480

QY 1342 GTCAATCAATATATAGATGACAAAGAGCCCTTAAGAGTGAACAGTACTC 1401  
DB 481 GGGAAATCATTTCTGAGACATTTCCCTCAAGAAAGAGAGAGAACCCCAAAATGT 540

QY 1402 AGTTGATACTGGAAGACCGGCCCCCAAAAGCTGATTTCCAGACCATGATCCAT 1461  
DB 541 GCTGATGCTTAAGGAGCTGGATGACAGAGGTTCTTACACAGACCATGATCAT 600

QY 1462 CCAG-----GGAGAGCTGACGCGGTGATACATGAGACATGATATGCTG 1515  
DB 601 CCAGGCTCCAGGGAATCTGAGATGGATATCAACAGAGAACCCAGACACAAATGGAGG 660

QY 1516 CATCCCTGCCCGAATATGAAGGGGAAAGGCTTCTGCTTATATTTGGCATCATGATCAT 1575  
DB 661 CATTCACAGCTTAAGAGCATAGGAGGAGAAACCTATTTATTTATTTAGGACATTTATGACAT 720

QY 1576 TCTACAGCTTACAGCTTTTGAAGATTGAGACATCTTTGAAAGCCCTGATACATGA 1635  
DB 721 TCTGATCATATATAGTTAATGAAGATTGAGACATTTCCGGAAGCTCTGTTTATGA 780

QY 1636 CGGAGACATGCTCAATGATGAGCCCAAGGCTTCTTACGTTGACGCTTCAACGCTTCA 1695  
DB 781 TGGGACACATGCTTCTTATATAGCAAGCTTTATGACAGACAGATTTCTTAAGTTTCA 840

QY 1696 GTGCAACAGCTATTTAAGAAAGTCT---CCTTGAAGCTTCTCTCCAAAA 1746  
DB 841 GAATTCAGAGTTCCTCAAGAAATTCAGGCTTGAAGGCTTCAACGCTTAAGAA 894

RESULT 11  
ABL65517  
ID ABL65517 standard; DNA: 1899 BP.  
AC ABL65517;  
XX  
XX  
DT 15-MAY-2002 (first entry)  
XX

DE Lung cancer related gene sequence SEQ ID NO:3854.  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
XX stomach; lung; prostate; pancreas; carcinoma; antileukemia; cancerous;  
KW cytostatic; gene therapy; antineoplastic; wilm's tumour; adenocarcinoma;  
XX gene; ds.  
OS Homo sapiens.  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US10838.  
XX  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 23-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 28-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
PA (AVAL-) AVALON PHARM.  
XX  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
PS Claim 1; SEQ ID 3854; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical

CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (1) of a signature gene set, where (1)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL1664  
 CC to ABL10110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (1) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.  
 XX  
 SQ Sequence 1899 BP; 601 A; 370 C; 377 G; 551 T; 0 other;

Query Match 11.0%; Score 408; DB 24; Length 1899;  
 Best Local Similarity 67.7%; Pred. No. 3e-104;  
 Matches 605; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

OY 862 TGAGCGCGGATGAACTGTAGCTGAGCTGAGTGTGCTCCATCTATGTGCGAG 921  
 DB 1 TGAACCTCTAATAGAACTGTCTAACCCTGAGCCAGTGATCTGTTTGTGTACAG 60  
 OY 922 CGACGATGAGTTCATTATTAAGACAGTCCATTAAGAGGCGGAATTTCTCAGAGCT 981  
 DB 61 TGATGATGAATTTATCATCAAAAGCTTCAGCAAAAGAGCTGAGTTCTTCAGAGCT 120  
 OY 982 GCTTCAGAGATACATGAACTCAACCCAGAACCCCTGCGACTTGTGCTTAATTCGA 1041  
 DB 121 ACTGCCAGGCTATACATGAATTTAAACCAGATCCAAAGACTCTTTGCCAAATTTTA 180  
 OY 1042 TGGACTGACTGTGCGAGCGAGTGGCAAGAACTGCGATTTGCTATGATCAATCT 1101  
 DB 181 CGGACTGTATGTATGACATGACAGAGGCAATTAATACAGATTTGTGTATGAACACCT 240  
 OY 1102 TTTACCAAGATCGGTAAATATCATATCATATGACCTCAAGGCTCAACTACAAAG 1161  
 DB 241 TTTGCGAGCTTCATGAGATGACATTTATGATGATTTGAAGGCTCAACGATATAGG 300  
 OY 1162 GCGGCGTTCAGAGAGAGAGAGAGAGAGCTTCCACATTTAAAGACCTAGACTTCT 1221  
 DB 301 AAGAGCATTCGCGTAAAGAGAGAGAGAAATCCAAACCATTTAAGACTTATGATTTCT 360  
 OY 1222 ACAAGACATCCCTGATGCTTTTGTGATGCTGACATGTCACAGCTCTCTGTAAGC 1281  
 DB 361 GCAGACATGACAGAGAGGTTGTATTTGATGAGAAACATCAACGCGCTTATGAAAGC 420  
 OY 1282 CCTGCGAGGTTGACTTTGCTGCTCAGAGCTTCAAGATATGATTAAGCTCTTGAT 1341  
 DB 421 ACTTGAGAGAGAGCTCGGCTGCTGAAAGCTTCAAGATCATGANTATAGCTTCTGT 480  
 OY 1342 GTCAATTCATATATGATCATGCAACAGAGAGCCCTTAAGCAGTGAACACAGTACTC 1401  
 DB 481 GGGAAATTAATTCCTGAGCAATTCCTCCAAAGAAAGAGAGAGAGAGAGAGAGAGT 540  
 OY 1402 AGTTGATCTCGAAGACGCGCCGCCCAAAAGGCTGTCTTCCAGACCATGGAATCCAT 1461  
 DB 541 GCCTATGCTAAGCGGACGCGATGCGATGACAGAGGTTCTCTACTCAACGCGCATGGAATCTAT 600  
 OY 1462 CCAG-----GGAGAGGCTCGACGGGCTGTACCATGAGAGCTGATGACCATATGGGTGG 1515  
 DB 601 CCAGGCTCCAGGGAATCTCGAGATGGGATATACAGAGAAACCCAGACACATAGGAGG 660  
 OY 1516 CATCCCTCCCGGAATAGTAAAGGGAAGGCTTCTGCTTATATATGTCATATGACAT 1575  
 DB 661 CATTCAGGCTAAAGCAATAGGGAAGAAATCTACTTTATTTATGCGCATATATTGACAT 720  
 OY 1576 TCTACAGCTTACAGGTTTGTAAAGAGTGGAGACCTTGGAAAGCCCTGGTACATGA 1635  
 DB 721 TCTGCAATCATATAGGTTATATGAAGAGATTCCTCTGGAAGGCTTGTGTTATGA 780

OY 1636 CGAGACACTGTCTCATGTCATGCCCCAGGCTTCTACGCTGAACGGTTCACGCTTCAT 1695  
 DB 781 TGGGACACTGTTCTGTTTCATATAGACCAAGCTTTATGACAGACATTTCTTAAGTTAT 840  
 OY 1696 GTGCAACACAGTATTTTAAGACATTC---CTTGAAGCTTCTCTCCCAAAA 1746  
 DB 841 GAATTCAGAGGTTTTCAGAAAATTCAGCTTTGAAGGCTTCACGCTTAAGAA 894

RESULT 12  
 ABL06531  
 ID ABL06531 standard; cDNA; 1795 BP.  
 XX  
 AC ABL06531;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14075.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 XX  
 DR P-PSDB: ABB62428.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1: SEQ ID NO 14075; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pt\_sequences.  
 XX  
 SQ Sequence 1795 BP; 467 A; 509 C; 477 G; 342 T; 0 other;

Query Match 9.6%; Score 357.4; DB 23; Length 1795;  
 Best Local Similarity 58.4%; Pred. No. 5.9e-90;  
 Matches 733; Conservative 0; Mismatches 441; Indels 81; Gaps 3;

OY 548 AAAATATGCGCATATGAGTGTGATTCCTCAGAGACACATATATAAAGACACACTCA 607  
 DB 411 AAGATTGGCCACAGAGCGCTCGGCGAGGCGGAGATTAACGTAACAAAAGATCCAGAG 470  
 OY 608 TCAGCCTTAAGAGTGCATCCAGTTAGCATTAACCAACACTGTGGGAGCCGTGAGTACC 667  
 DB 471 TCGCAGATCATGCGGCTCATCCAACTGCGAATTCACACACACTGTGCGAGCTGCGCTCG 530



Db 971 ACGATTTCGTTCCGCCAGATGTTCCATACACCTGCCCATCTACACAGCATTC 1030  
 Qy 773 GGTTCAGACCTATGACATCTGTTGCTTCCGCTACTTCCGGAGCTATTTGATCCGG 832  
 Db 1031 CGCTTCAGAGCTCTACGACCCATGCTTCTGCTTCCGAGATCTGTGGATTCGA 1090  
 Qy 833 CCGGATGATTAATGTTCCCTCTGAGTGAGCCGCTGATGAACTCTGATGCTGGA 892  
 Db 1091 CCGGATGATTAATGTTCCCTCTGAGTGAGCCGCTGATGAACTCTGATGCTGGA 1150  
 Qy 893 GCTAGTGGTCCCTATGATGTCGACGAGTGTGATGATGATGATGATGATGATGAT 952  
 Db 1151 GCTAGTGGTCCCTATGATGTCGACGAGTGTGATGATGATGATGATGATGATGAT 1210  
 Qy 953 CATAAAGAGGCGAATTTCTGAGAGCTGCTTCCAGATATACATGATGATGATGATGAT 1012  
 Db 1211 AAGAAAGAAAGAGATTTCTGAGAGCTGCTTCCAGATATACATGATGATGATGATGAT 1270  
 Qy 1013 AACCTCGGACTTCTGCTGCTTAAATTTATGAGATGATGATGATGATGATGATGAT 1072  
 Db 1271 AATCTCGAAGCTGCTGCTTAAATTTATGAGATGATGATGATGATGATGATGAT 1330  
 Qy 1073 AACATGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1132  
 Db 1331 AACGTCGGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1390  
 Qy 1133 TATGACCTCAAGGCTCAACCTACAAAGCGGCTTCCAGAAAGAGAGAGAGAGCT 1192  
 Db 1391 TACGACCTTAAAGGATCTTCTCTCGACGAGAGAGCTTCCAGAGAGAGAGAGAGCT 1450  
 Qy 1193 CTTCACATTTTAAAGCTTACCTTCTTACAAAGACATCCCTGATGATGATGATGATGAT 1252  
 Db 1451 AGTCCACGTTCAAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1510  
 Qy 1253 GCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312  
 Db 1511 ACGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1570  
 Qy 1313 TTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1365  
 Db 1571 TTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1623  
 RESULT 14  
 ABL15962/c  
 ID ABL15962 standard; cDNA; 6699 BP.  
 XX  
 AC ABL15962;  
 XX  
 AC 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42368.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN MO200171042-A2.  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB71859.

XX  
 PR New isolated nucleic acid detection reagent for detecting 1000 or more  
 PR genes from Drosophila and for elucidating cell signalling and cell-cell  
 PR interactions -  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 42368; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB13511), expressed DNA  
 CC sequences (AB16173-AB16175) and the encoded proteins  
 CC (AB16173-AB16175).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcr\_sequences.  
 XX  
 SO Sequence 6699 BP; 1663 A; 1551 C; 1541 G; 1944 T; 0 other;  
 Query Match 9.3%; Score 346.6; DB 23; Length 6699;  
 Best Local Similarity 63.5%; Pred. No. 1.5e-86;  
 Matches 529; Conservative 0; Mismatches 304; Indels 0; Gaps 0;  
 Qy 533 GGCATGCCCATCAAGAAATAGCCATAGAAAGTGTGATTCCTCAGAGAGACAATAT 592  
 Db 3519 GGGAAAGCAGAAAG 3460  
 Qy 593 AAAAAGACAACTCATGACCTTGAAGAGTGCATTCAGATGAGATGAGATGAGATGAG 652  
 Db 3459 AAGAAGATCAATCAAGCAAAATAGGCTCATTCAGCTGGATCCAGATGAGATGAG 3400  
 Qy 653 GGGAGCTGAGTACCAAG 712  
 Db 3399 GGCAGCCTGCGCAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3340  
 Qy 713 AGTATCTCTTCCAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 772  
 Db 3339 ACGATTTGCTTCCGAG 3280  
 Qy 773 GGTTCAGAGACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 832  
 Db 3279 CGCTTCAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3220  
 Qy 833 CCGGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 892  
 Db 3219 CCGGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3160  
 Qy 893 GCTAGTGGTCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 952  
 Db 3159 GCTAGTGGTCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3100  
 Qy 953 CATAAAGAGGCGAATTTCTGAGAGCTGCTTCCAGATATACATGATGATGATGATGAT 1012  
 Db 3099 AAGAAGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3040  
 Qy 1013 AACCTCGGACTTCTGCTGCTTAAATTTATGAGATGATGATGATGATGATGATGATGAT 1072  
 Db 3039 AATCTCGAAGCTGCTGCTTAAATTTATGAGATGATGATGATGATGATGATGATGAT 2980  
 Qy 1073 AACATGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1132  
 Db 2979 AACGTCGGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2920  
 Qy 1133 TATGACCTCAAGGCTCAACCTACAAAGCGGCTTCCAGAAAGAGAGAGAGAGCT 1192  
 Db 2919 TACGACCTTAAAGGATCTTCTCTCGACGAGAGAGCTTCCAGAGAGAGAGAGAGCT 2860  
 Qy 1193 CTTCACATTTTAAAGCTTACCTTCTTACAAAGACATCCCTGATGATGATGATGATGAT 1252  
 Db 2859 AGTCCACGTTCAAGGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2800

```
QY 1253 GCTGACATGACAAAGCTCTCTCTGTAAAGCCCTGACGCTGCTGTTGGTGTGACAGAC 1312
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2799 ACGGACACAGATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 2740
QY 1313 TTCAAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTC 1365
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2739 TTTCAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTC 2687

RESULT 15
AAH73319
ID AAH73319 standard; cDNA; 674 BP.
XX
AC AAH73319;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 4593.
XX
KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WC200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.

(MILL-) MILENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1; Page 1021; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 674 BP; 165 A; 163 C; 165 G; 180 T; 1 other;

Query Match 8.7%; Score 324.8; DB 22; Length 674;
Best Local Similarity 93.9%; Pred. No. 5.4e-81;
Matches 338; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1377 CCTTAAGCAGTAAACACAGTACTCTAGTGTGATCTGGAAGACCGGCCCCCAAAAGGCTC 1436
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 CCTTAAGCAGTAAACACAGTACTCTAGTGTGATCTGGAAGACCGGCCCCCAAAAGGCTC 61
QY 1437 TGTATTCACAGCCATGAAATTCATCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 TGTATTCACAGCCATGAAATTCATCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
QY 1497 CTGATGACCATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 1556
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 122 CTGATGACCATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 181
```

```
QY 1557 ATATTGGATCATTTGACATTTCTACAGTCTTACAGGTTTGTAAAGATTGAGACACTCTT 1616
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 ATATTGGATCATTTGACATTTCTACAGTCTTACAGGTTTGTAAAGATTGAGACACTCTT 241
QY 1617 GGAAGCCCTGTGATGACAGAGACACTGTCTGATGATGATGATGATGATGATGATGATG 1676
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 242 GGAAGCCCTGTGATGACAGAGACACTGTCTGATGATGATGATGATGATGATGATGATG 301
QY 1677 AACGTTCCAGCGCTTCATGATGACACAGATTAATTAAGAGATTCCCTTGAAGCCTTCTC 1736
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 302 AACGTTCCAGCGCTTCATGATGACACAGATTAATTAAGAGATTCCCTTGAAGCCTTCTC 361
```

Search completed: January 8, 2003, 14:08:34  
Job time : 746 secs



QY	1767	TCCTCGCGGAGAGAGGCTCCAGTAGGGCAACTCTGCATTACTTACAGAGCAGCAGGCTCTGG	1822
Db	553	TCCTCGCGGAGAGAGGCTCCAGTAGGGCAACTCTGCATTACTTACAGAGCAGCAGGCTCTGG	494
QY	1827	GGGAAACAAGGCAACAAGTACACAACAAGGACAGATGGAGCCAGGCGTTACCTTGGTC	1886
Db	493	GGGAAACAAGGCAACAAGTACACAACAAGGACAGATGGAGCCAGGCGTTACCTTGGTC	434
QY	1887	GTCCTGATGTTTAACTCAACATCTCCACTTTGGAGAAATCAAGTAGGGCTGGCTATTTC	1946
Db	433	GTCCTGATGTTTAACTCAACATCTCCACTTTGGAGAAATCAAGTAGGGCTGGCTATTTC	374
QY	1947	CTGACCCCAAGTTTCTACGCTCTAGTTGGAGAGACTTTGGAATGCAATCAACTACAA	2006
Db	373	CTGACCCCAAGTTTCTACGCTCTAGTTGGAGAGACTTTGGAATGCAATCAACTACAA	314

Oy	2007	CCTTGAAAGGTTGAAGTGGACATGCAGAGTCACCATTTAAGGCCAAGGCGTCA		2066
Dd	313	CCTTGAAAAGGTGAAGTAGCAGAGTCCAGATTACCATTAAGGCCAAGGCGTCA		254
Oy	2067	GACCTGGAACAAGAATYTCGSCACTCTCTGTGATCCCAAGATGTCAGCCCTTGCCCCACAA		2126
Dd	253	GACCTGGAACAAGAATYTCGSCACTCTCTGTGATCCCAAGATGTCAGCCCTTGCCCCACAA		194
Oy	2127	TGCTCAATTTCTTCTACTGTGTCATCAAAAAAGAGATGTAATAAGTAGAGGAGCTG		2186
Oy	2187	CTCTCCATCTTCTTCTCCTGAGAACCTTCTCTCTCTCTCTCTCTCAATGAATGGCC		2246
Dd	133	CTCTCCATCTTCTTCTCCTGAGAACCTTCTCTCTCTCTCTCTCAATGAATGGCC		74
Oy	2247	TTAGTGCCCTCAGAGATTTGAGACCCGACAGATCCCCTCCACTCACAGTTGGGTGAC		2306
Dd	73	TTAGTGCCCTCAGAGATTTGAGACCCGACAGATCCCCTCCACTCACAGTTGGGTGAC		14
Oy	2307	GATTTCAACTGG 2319       		
Dd	13	GATTTCAACTGG 1		
RESULT 2				
US-09-864-761-13058 ; Sequence 13058, Application US/09864761				
; Patent No. US20020048763A1				
GENERAL INFORMATION:				
APPLICANT: Penn, Sharon G.				
APPICANT: Rank, David R.				
APPLICANT: Hanzel, David K.				
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR				
FILE REFERENCE: Aecmlca-X-1				
CURRENT APPLICATION NUMBER: US/09/864,761				
CURRENT FILING DATE: 2001-05-23				
PRIOR APPLICATION NUMBER: US 60/180,312				
PRIOR FILING DATE: 2000-02-04				
PRIOR APPLICATION NUMBER: US 60/207,456				
PRIOR FILING DATE: 2000-05-26				
PRIOR APPLICATION NUMBER: US 09/632,366				
PRIOR FILING DATE: 2000-08-03				
PRIOR APPLICATION NUMBER: GB 24263.6				
PRIOR FILING DATE: 2000-10-04				
PRIOR APPLICATION NUMBER: US 60/236,359				
PRIOR FILING DATE: 2000-09-27				
PRIOR APPLICATION NUMBER: PCT/US01/00666				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00667				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00664				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00669				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00665				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00668				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00663				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00662				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00661				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: PCT/US01/00670				
PRIOR FILING DATE: 2001-01-30				
PRIOR APPLICATION NUMBER: US 60/234,687				
PRIOR FILING DATE: 2000-09-21				
PRIOR APPLICATION NUMBER: US 09/608,408				
PRIOR FILING DATE: 2000-06-30				

```

PRIORITY APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ. ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 13058
LENGTH: 598
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL096800.20
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.53
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65
US-09-864-761-13058

Query Match          12.7%  Score 469.8;  DB 10;  Length 598;
Best Local Similarity 90.1%  Pred. No. 2e-119;
Matches 527;  Conservative 0;  Mismatches 52;  Indels 6;  Gaps 2;

QY  1331  AGCCCTTGAGTGAATCCATATATATAGATCATGCACACAGAGAGCCCTTAAGCAGTGA 1390
DB  2      AGCCCTGCGGTGCATCCACATATATAGATCATGCACACAGAGAGCCCTTAAGCAGCAG 61
QY  1391  AC---ACAGTACTAGTTGATACCTCGAAGACCGGCCCCCAAAAGGCTCTGTATCCACA 1447
DB  62      ACTCTTCAAGTGTCAATGCATGCACCTCAAGACTGGCTCCCAAAAGGCTCTGTATCCACA 121
QY  1448  GCCATGGAATCATCCACAGAGAGAGGCTCGACGGGGTGTATCCATGAGACTGATGACCAT 1507
DB  122     GCCATGGAATCATCCACAGAGAGAGGCTCGGCTGGGCGACACATGAGAGCCGATGACCAT 181
QY  1508  ATGGGTGGCATCCCTGCCCGGAATGTAAGGGGAAGGCTTGTCTTATATATGGCATC 1567
DB  182     AAGGGTGGCATCCCTGCCCGGAATGTAAGGGGAAGGCTTGTCTTATATATGGCATC 241
QY  1568  ATTGACATTTCACATCTTACAGGTTTGTAAAGAGTTGGAGCAGCTTTGGAAAGCCCTG 1627
DB  242     ATTGACATTTCACATCTTACAGGTTTGTAAAGAGTTGGAGCAGCTTTGGAAAGCCCTG 301
QY  1628  GTACATGACGAGACACGTCTCTACGTGCATGCCAGGCTTCTACGCTGAACGGTTCCAG 1687
DB  302     GTACATGATGGGGAGCGCTGTCTACGTGCATGCCAGGCTTCTACGCTGAACGGTTCCAG 361
QY  1688  GCCTTCAGTGCACACAGATATTTAAGAAAGATTCCTTGAAGCCCTTCCCTCCAAAAAG 1747
DB  362     CACTTCATGTGCAACGACAGTATTTAAGAAAGATTCCTTGAAGCCCTTCCCTCCAAAAAG 421
QY  1748  TTTCGGTCTGAGCTATCTTCTCTCGGCGAGCAGGCTCCAGTGGCACTCCTGCATTACT 1807
DB  422     TTTCGGTCTGAGCTATCTTCTCTCTCGTCTGATACGGGCTCCAGTGGCACTCCTGCATTACT 481
QY  1808  TACCAAGCATCGCTCTCGGGGGAACAACAAGGCACAAAGTGAACAACAAGGAGGTGGAG 1867
DB  482     TACCAAGCATCGCTCTCGGGGGAACAACAAGTGCACAAAGTGAACAACAAGGAGGTGGAG 538
QY  1868  CCAGCGGTTTCACCTTGTCGTCTGATGTTTATACCTCAGACTCCA 1912
DB  539     CCAGGTTTCACCTTGTCGTCTGATGTTTATACCTCAGACTCCA 583

RESULT 3
US-09-954-456-827
; Sequence 827, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617

```



QY 3643 GTTCAGAAATGCAATATACATATAAAGTATATGTTTAATGTAATAACTTTAA 3702  
DB 72 GTTCAGAAATGCAATATACATATAAAGTATATGTTTAATGTAATAACTTTAA 13  
QY 3703 TGAGTTATTTTA 3713  
DB 12 TGAGTTATTTTA 2  
RESULT 5  
US-09-864-761-29620  
Sequence 29620, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmlca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 29620  
LENGTH: 346  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL096800.20  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.53  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65  
OTHER INFORMATION: SWISSPROT HIT: P38994, EVALUATE 4.00e-07

OTHER INFORMATION: NT HIT: g14505814, EVALUATE 1.00e-125  
OTHER INFORMATION: EST\_HUMAN HIT: BE794576.1, EVALUATE 1.00e-125  
US-09-864-761-29620  
Query Match 7.7%; Score 285; DB 10; Length 346;  
Best Local Similarity 92.3%; Fred. No. 1.3e-68;  
Matches 300; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 1400 TCAGTTGATACATCGAAGACCGCCGCCCAAGGCTGTATTCACAGCCATGAGATCC 1459  
DB 22 TCATATGACACTCAAGAGACTGGCTCCCAAGGCTGTATTCACAGCCATGAGATTC 81  
QY 1460 ATCCAGGAGAGAGCTCGACGCGGTGTACCATGAGACTGATGACCATATGGGTGCATC 1519  
DB 82 ATCCAGGAGAGAGCTCGCGCTGGGCGCACCATGAGGCGCATGACCATATGGGTGCATC 141  
QY 1520 CCTGCCCGGAATAGTAAGGAGGAAAGGCTTCGCTTATATGGCATATTCATTCATCA 1579  
DB 142 CCTGCTCAGAAATAGTAAGGAGGAAAGGCTTCGCTTATATGGCATATTCATTCATCA 201  
QY 1580 CAGTCTTACAGGTTTGTATTAAGAGTTGAGACACTCTTGGAAAGCCCTGTACATGACGGA 1639  
DB 202 CAGTCTTACAGGTTTGTATTAAGAGTTGAGACACTCTTGGAAAGCCCTGTACATGATGGG 261  
QY 1640 GACACTGTCTCAGTGCATCGCCAGGCTTCTACGCTGAACGCTTCACGCTTCATGTGC 1699  
DB 262 GACGCTGTCTCAGTGCATCGCCAGGCTTCTACGCTGAACGCTTCACGCTTCATGTGC 321  
QY 1700 AACACAGTATTTAAGAAGATTCCT 1724  
DB 322 AACGCAGTATTTAAGAAGATTCCT 346  
RESULT 6  
US-09-783-590-9100  
Sequence 9100, Application US/09783590  
Patent No. US20020110850A1  
GENERAL INFORMATION:  
APPLICANT: Dillon, Patrick J.  
APPLICANT: Haseltine, William A.  
APPLICANT: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
FILE REFERENCE: PO-16,2C1  
CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9100  
LENGTH: 350  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (24)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (40)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (44)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (134)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (172)  
OTHER INFORMATION: n equals a,t,g, or c

```

1 NAME/KEY: misc feature
2 LOCATION: (173)
3 OTHER INFORMATION: n equals a,t,g, or c
4 NAME/KEY: misc feature
5 LOCATION: (184)
6 OTHER INFORMATION: n equals a,t,g, or c
7 NAME/KEY: misc feature
8 LOCATION: (189)
9 OTHER INFORMATION: n equals a,t,g, or c
10 NAME/KEY: misc feature
11 LOCATION: (193)
12 OTHER INFORMATION: n equals a,t,g, or c
13 NAME/KEY: misc feature
14 LOCATION: (203)
15 OTHER INFORMATION: n equals a,t,g, or c
16 NAME/KEY: misc feature
17 LOCATION: (209)
18 OTHER INFORMATION: n equals a,t,g, or c
19 NAME/KEY: misc feature
20 LOCATION: (217)
21 OTHER INFORMATION: n equals a,t,g, or c
22 NAME/KEY: misc feature
23 LOCATION: (245)
24 OTHER INFORMATION: n equals a,t,g, or c
25 NAME/KEY: misc feature
26 LOCATION: (250)
27 OTHER INFORMATION: n equals a,t,g, or c
28 NAME/KEY: misc feature
29 LOCATION: (256)
30 OTHER INFORMATION: n equals a,t,g, or c
31 NAME/KEY: misc feature
32 LOCATION: (282)
33 OTHER INFORMATION: n equals a,t,g, or c
34 NAME/KEY: misc feature
35 LOCATION: (298)
36 OTHER INFORMATION: n equals a,t,g, or c
37 NAME/KEY: misc feature
38 LOCATION: (317)
39 OTHER INFORMATION: n equals a,t,g, or c
40 NAME/KEY: misc feature
41 LOCATION: (321)
42 OTHER INFORMATION: n equals a,t,g, or c
43 NAME/KEY: misc feature
44 LOCATION: (323)
45 OTHER INFORMATION: n equals a,t,g, or c
46 NAME/KEY: misc feature
47 LOCATION: (350)
48 OTHER INFORMATION: n equals a,t,g, or c
49
50 US-09-783-590-9100

```

Query Match	4.1%	Score 152	DB 10	Length 350
Best Local Similarity	69.9%	Pred. No. 5.9e-32		
Matches 232	Conservative 0	Mismatches 95	Indels 5	Gaps 3
QY 2448	GGACAGACTAGCTGGCACATTATCCCTACCTAGTCTCTCTCTCTCGAGCT-CCCTGGAGA	2506		
DB 9	GAACAGACTAGCTGGCMACATTATCCCAACCTNAGTNCCTCTCTGATCCCTGGAGAA	68		
QY 2507	ATACCTCTGTAATCTCTCTGTAAGAATTTTTGGGGGATAAGGAGTGTAAACACCTCCAGC	2566		
DB 69	AAACCCCTGATATCTCTGACAGAGTTTTGGGGGATAAGGAGTGTAAACACCTGCCCA	128		
QY 2567	TTTCTCTCTCTTTTTTTTTTCTGAAAAAAGAAAAACACAGCAGACACATTTCAAGCC	2626		
DB 129	CTT-TGNTTTTTTTTTTCTCCGAAAAAAGGGGAAAAAGGACACAGNNCCATTTTAAAGCC	187		
QY 2627	ATTTCAGATCCGAACCTCCAGAAGTGTGACAGAAGTGCCTATTTCGTAGAGTTCCCTCAGA	2686		
DB 188	ATTTTNAGTTCAAAANTCCAGAGGTTTTGNGGGGTTCCTTCCGTGGGATTCCTCCCTNAG	247		
QY 2687	AGAG---CCATGGCTTTATGAAGAAGAGATGATTCGCTCTCCAGAAACAGAGTCTT	2743		
DB 248	AANGGCCCGGGTTTTTTTTAAAGGGGAAAAAGGTGGNCTTTGCTCTCGCAANAGGATTTCC	307		

```

OY      2744  CTTTAACTCCTCCTCCTCTGATGAAATTTCTT  2775
          ||| ||| | | | | | | | | | |
Db      308  TTTTAAATNCCCNANTTTTGGGGGAATTTTT  339

```

RESULT 7  
US-09-938-842A-2662  
; Sequence 2662, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:

```

1  APPLICANT: Harper, Jeff
2  APPLICANT: Krebs, Joel
3  APPLICANT: Wang, Xun
4  APPLICANT: Zhu, Tong
5  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
6  TITLE OF INVENTION: SAME, AND METHODS OF USE
7  FILE REFERENCE: SCRIPT300-3
8  CURRENT APPLICATION NUMBER: US/09/938, 842A
9  CURRENT FILING DATE: 2001-08-24
10 PRIOR APPLICATION NUMBER: US 60/227, 866
11 PRIOR FILING DATE: 2000-08-24
12 PRIOR APPLICATION NUMBER: US 60/264, 647
13 PRIOR FILING DATE: 2001-01-16
14 PRIOR APPLICATION NUMBER: US 60/300, 111
15 PRIOR FILING DATE: 2001-06-22
16 NUMBER OF SEQ ID NOS: 5379
17 SEQ ID NO 2662
18 LENGTH: 2580
19 TYPE: DNA
20 ORGANISM: Arabidopsis thaliana
21 US-09-938-842A-2662

```

Query Match	2.0%;	Score 75;	DB 9;	Length 2580;
Best Local Similarity	49.2%;	Pred. No. 3.7e-10;		
Matches 325;	Conservative	0;	Mismatches 315;	Indels 21;
				Gaps 4;

Oy	716	ATCTTCTTCTCCACGAGGAGCAACCTGACCCCTGCTCATTCACATGACTTTCGT	775
Db	1426	ATGTCTCTCCCTAGAGACGGCTCCAACTTTTACCCCTCAGACACAATCTGTGCACTTCTCC	1485

**OY**

776 TTCAAGACCTATGCACCTGTTCGCGTCCGGACTATTGTATC GGCCC 835  
| | |||| | | | | | | | |  
| | |||| | | | | | | | |

DB 1480 TGGAAAGACATATGTCCTCCAGATGGTTTCAGGAATTAAAGCAGATGTTCAAGTTAGATCGT 1543

QY 836 GATGATTAATCTGTATTCCCTCTGCAGTGAAGCG---CTGATTGAATCTGTAGCTCTGGA 892

Db 1546 GCAGGATACATGATGTCCTATTGTGTGATGATGAGCCCTGACAGAAATTTCTTCCCCCGG 1605

```

      |||||  - |||||  - ||  |||||  - ||  |||||  - ||
Db  1606 AAAAGTGGCAGTATCTTCTACCTTCTCATGACGACAGATTTGTGATCAAGACTTTAAAA 1665

```

Qy 953 CATTAAGAGCGCGAATTCTCTCAGAAAGCTGCTTCCAGGATACTACATGAAACCCITCAACCG 1012

Db 1666 AAGTCTGAGTTCGAGTTCCTACTCAGATGTTGCCCTAAGTACTATGAACATGTAGGCGAG 1725

QY 1013 AACCTGGACTTTCGCTAAATTCATAGGACTGACTGTGT--GCAGCGAGTGGC 10699

1070 AAGAACATTCGGATCTGTGATGAACAATCTTTACCAAGATCGGTAAAAATGCATATTC 1129

Db 1786 AAAAGGTACGCTTGTAGTCATGGGAAATATGTTTGCACAGAAATTGAAGATTCATCGT 1845

Db 1846 CGTTATGATCTAAAGGTTCAAC--TCAGGAGATTACTGAAAGATCAAAATCCAA 1902

```

QY  1190 CCCTTCCCATTTAAAGACTAGACTTCTACAAACATCCCTGATGGCTTTTGTG 1249
      ||||| |||||  || || ||||  || ||
Db  1903 GAGAAGACCACTTGAAGATCT-----TCATCTAGCTTATGAATTCATATG 1950

```

QY 1250 GATGCTGACATGTACACACGCTCTCTGTAGAACCCGTGCAGCGCTGACTGTGGTCTGCAG 1309  
DB 1951 GACAAAGCTGTGTACGGGAGCCCTTTTCACAAATTTTACTTACAGCTCTGCTTTTGGAA 2010  
QY 1310 AGCTTCAGATTAATGATTTACAGCGCTCTGTATGATTCATTCATTAATATAGATGACAA 1369  
DB 2011 TCGGTGAACATTCATGTACTGACAGCTCTTTTACTGGGATTCATTTTACAGCTCTCGGCCAA 2070  
QY 1370 C 1370  
DB 2071 C 2071

## RESULT 8

US-10-012-055-3  
; Sequence 3, Application US/10012055  
; Patent No. US20020164750A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel A.  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 56634, A NOVEL HUMAN PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE  
; FILE REFERENCE: 10448-114001  
; CURRENT APPLICATION NUMBER: US/10/012,055  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/248,325  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1266  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-012-055-3

Query Match 2.0%; Score 73.8; DB 9; Length 1266;  
Best Local Similarity 48.8%; Pred. No. 5e-10;  
Matches 297; Conservative 0; Mismatches 297; Indels 15; Gaps 3;

QY 756 ATCTACATCAATGACTTTCGTTCAAGACCTATGACCTGTGCTTCCGCTACTTCCGGG 815  
DB 278 ATCTGCCACTCATTTTCATGACAGAGATTTGTCGCCAGTCTTCAGGAACCTCCGCTG 337  
QY 816 AGCTATTTGGTATCGGCGCCGATGATTAATTAATTCCTCTGACAGTACCGCTGATTTG 875  
DB 338 ATGATTTGGCATTTGATGACCAAGATTAATTAATTCCTCTGACAGTACCGCTGATTTG 397  
QY 876 AACTCTGTACTCTGAGCTAGTGTGCTCTTCTATGATGATGATGATGATGATGATGATG 935  
DB 398 AAGG-----TGAAGGAGTGTGTGTGCTCTTCTATGATGATGATGATGATGATGATG 448  
QY 936 TTTTAAAGACATTCACATTAAGAGCGGAAATTTCTGCAAGAGCTGCTTCCAGATTA 995  
DB 449 TCATCAAGAGATTCACATTAAGAGCGGAAATTTCTGCAAGAGCTGCTTCCAGATTA 508  
QY 996 ACATGAACCTCAACAGAACCTCG---GACTTTGCTGCTTAATTTCTATGAGCTGACT 1052  
DB 509 ACCAGTACATTTGAAAGTGTGATGCAACGCGTTCTGCCCCAGTTCTGCGGAGATGATCC 568  
QY 1053 GTGTGAGCAGAGTGTGCAAGAACATTCGATTTGTGTGATGATGATGATGATGATGATG 1112  
DB 569 GATTCAGTGTGACACAG 628  
QY 1113 CGGTAAATATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1172  
DB 629 GTCTTCTGTGACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 688  
QY 1173 AGAAAG 1232  
DB 689 ATTAAG 748  
QY 1233 CTGATGCTCTTTTGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1292

DB 749 AGAAAGATATATATTTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805  
QY 1293 ACTGTTGTGTGTCAGAGAGCTCAAGATTAATGATTAATGATTAATGATTAATGATTAATG 1352  
DB 806 ATGAGAGATTTAGTGTGACACTGAAGATTAATGATTAATGATTAATGATTAATGATTAATG 865  
QY 1353 ATATGATC 1361  
DB 866 ACATCATTC 874

## RESULT 9

US-10-012-055-1  
; Sequence 1, Application US/10012055  
; Patent No. US20020164750A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel A.  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 56634, A NOVEL HUMAN PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE  
; FILE REFERENCE: 10448-114001  
; CURRENT APPLICATION NUMBER: US/10/012,055  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/248,325  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3224  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (73)...(1335)  
US-10-012-055-1

Query Match 2.0%; Score 73.8; DB 9; Length 3224;  
Best Local Similarity 48.8%; Pred. No. 9.3e-10;  
Matches 297; Conservative 0; Mismatches 297; Indels 15; Gaps 3;

QY 756 ATCTACATCAATGACTTTCGTTCAAGACCTATGACCTGTGCTTCCGCTACTTCCGGG 815  
DB 350 ATCTGCCACTCATTTTCATGACAGAGATTTGTCGCCAGTCTTCAGGAACCTCCGCTG 409  
QY 816 AGCTATTTGGTATCGGCGCCGATGATTAATTAATTCCTCTGACAGTACCGCTGATTTG 875  
DB 410 ATGATTTGGCATTTGATGACCAAGATTAATTAATTCCTCTGACAGTACCGCTGATTTG 469  
QY 876 AACTCTGTACTCTGAGCTAGTGTGCTCTTCTATGATGATGATGATGATGATGATGATG 935  
DB 470 AAGG-----TGAAGGAGTGTGTGTGCTCTTCTATGATGATGATGATGATGATGATG 520  
QY 936 TTTTAAAGACATTCACATTAAGAGCGGAAATTTCTGCAAGAGCTGCTTCCAGATTA 995  
DB 521 TCATCAAGAGATTCACATTAAGAGCGGAAATTTCTGCAAGAGCTGCTTCCAGATTA 580  
QY 996 ACATGAACCTCAACAGAACCTCG---GACTTTGCTGCTTAATTTCTATGAGCTGACT 1052  
DB 581 ACCAGTACATTTGAAAGTGTGATGCAACGCGTTCTGCCCCAGTTCTGCGGAGATGATCC 640  
QY 1053 GTGTGAGCAGAGTGTGCAAGAACATTCGATTTGTGTGATGATGATGATGATGATGATG 1112  
DB 641 GATTCAGTGTGACACAG 700  
QY 1113 CGGTAAATATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1172  
DB 701 GTCTTCTGTGACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 760  
QY 1173 AGAAAG 1232  
DB 761 ATTAAG 820  
QY 1233 CTGATGCTCTTTTGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1292









**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 12:58:05 ; Search time 4885 Seconds  
(without alignments)  
12309.896 Million cell updates/sec

Title: US-10-003-354-3

Perfect score: 3713  
Sequence: 1 attaacagcgctggttagg.....aaacttaatgagttattta 3713

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hlc: \*  
9: gb\_estl: \*  
10: gb\_est2: \*  
11: gb\_hlc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vit: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rtd: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	859.4	23.1	889	14	BQ894466
2	841	22.7	1045	13	BM562816
3	828.2	22.3	1119	13	BM545034
4	827	22.3	865	9	AL522833
5	820.2	22.1	838	9	AU132382
6	817	22.0	1069	13	BM549810

c	7	813.6	21.9	824	9	AU126052
	8	802	21.6	814	9	AL528200
	9	800.6	21.6	904	14	BQ439604
	10	797.6	21.5	1085	13	BM562145
	11	779.2	21.0	1046	13	BM455340
	12	772.8	20.8	864	13	BI090667
	13	768.8	20.7	824	9	AU130897
	14	766	20.6	908	14	BQ930009
	15	760.4	20.5	783	9	AL522834
	16	754	20.3	876	14	BQ962454
c	17	721.4	19.4	882	9	AU149586
	18	715.8	19.3	875	13	BI181582
	19	694.2	18.7	919	14	BQ423398
	20	687.2	18.5	791	9	AL514476
	21	683.6	18.4	2105	11	BC020365
	22	680.4	18.3	2472	11	AK014403
	23	674	18.2	964	12	BE794576
	24	671.4	18.1	974	12	BG117080
	25	670.6	18.1	808	12	BG724319
	26	670.2	18.1	1043	13	BM461538
	27	665.4	17.9	705	9	AU130473
	28	660.4	17.8	803	12	BG679864
	29	654.4	17.6	683	13	BI862007
	30	653.4	17.6	685	9	AL710849
c	31	651.6	17.5	736	14	BM982986
	32	649.4	17.5	768	12	BG717502
c	33	648.4	17.5	682	14	BQ448271
	34	643.8	17.3	1182	12	BG119715
	35	638.8	17.2	737	12	BG392973
	36	629.6	17.0	671	13	BI259656
	37	616	16.6	683	13	BM015907
	38	615.2	16.6	973	12	BG025366
	39	614.8	16.6	1065	14	BQ242439
	40	609.2	16.4	629	14	BQ447995
	41	609.2	16.4	1108	13	BI117425
	42	607.6	16.4	780	10	BE535818
c	43	604	16.3	734	10	AM364156
	44	598.2	16.1	871	12	BF036836
	45	594	16.0	646	9	AL134930

## ALIGNMENTS

RESULT 1  
LOCUS BQ894466 889 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGENCOURT\_8727378 NIH\_MGC\_47 Homo sapiens CDNA clone IMAGE:6340661  
5', mRNA sequence.  
ACCESSION BQ894466  
VERSION BQ894466.1 GI:22286480  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 889)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
COMMENT Email: cga@pds-remail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory  
CDNA Sequencing by: The I.M.A.G.E. Consortium (ILNI)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNI at:  
http://image.llnl.gov  
Plate: LICM2537 row: h column: 06  
High quality sequence stop: 628.  
Location/Qualifiers  
1. 889

## FEATURES

source

Query Match	Best Local Similarity	Score	DB	Length
Matches 877; Conservative	98.7%; Pred. No. 2.9e-187;	859.4;	14;	889;
	0; Mismatches 11; Indels 1; Gaps 1;			
858	GCAGGAGCGCGTATGTAACCTGTAGCTGGAGCTAGTGGTCCCATCTATGCT	917		
Db	1 GCAGGAGCGCGTATGTAACCTGTAGCTGGAGCTAGTGGTCCCATCTATGCT	60		
918	CCAGGAGGATGATTCATTATTAAAGACAGTCCACATTAAGAGCGGAAATTTGCGAGA	977		
Db	61 CCAGGAGGATGATTCATTATTAAAGACAGTCCACATTAAGAGCGGAAATTTGCGAGA	120		
978	AGCTCTTCCAGGATACATCATGAAACCTTCACAGACCCCTGGACTTGTCTGCTAAAT	1037		
Db	121 AGCTCTTCCAGGATACATCATGAAACCTTCACAGACCCCTGGACTTGTCTGCTAAAT	180		
1038	TCTATGACTGTACTGTGTGCGAGGAGGTGGCAAAACATTGGAAATGGTGATGAAACA	1097		
Db	181 TCTATGACTGTACTGTGTGCGAGGAGGTGGCAAAACATTGGAAATGGTGATGAAACA	240		
1098	ATCTTTTACCAAGATCGGTAAATATGATATCAATATATGACTCAAAAGGCTCAACTACA	1157		
Db	241 ATCTTTTACCAAGATCGGTAAATATGATATCAATATATGACTCAAAAGGCTCAACTACA	300		
1158	AACGGCGGGCTTCCCGAGAAAGAGAGAGAAACCTCTTCCCACTTTTAAAGACTAGACT	1217		
Db	301 AACGGCGGGCTTCCCGAGAAAGAGAGAGAAACCTCTTCCCACTTTTAAAGACTAGACT	360		
1218	TCTTACAAGACATCCCTGATGCTCTTTTGGATGCTGACATGTACAACGCTCTCTGTA	1277		
Db	361 TCTTACAAGACATCCCTGATGCTCTTTTGGATGCTGACATGTACAACGCTCTCTGTA	420		
1278	AGACCTCGAGGATGATGTTGGTGGTGCAGACGCTTCAAGATATGATTAAGGCTCACT	1337		
Db	421 AGACCTCGAGGATGATGTTGGTGGTGCAGACGCTTCAAGATATGATTAAGGCTCACT	480		
1338	TGATGTCAATCATTAATATAGATCATGCAACAGAGAGCCCTTAAAGAGTGAACACAGT	1397		
Db	481 TGATGTCAATCATTAATATAGATCATGCAACAGAGAGCCCTTAAAGAGTGAACACAGT	540		
1398	ACTCAGTGTATCTGGAAGACCGCGCCGCCAAAGAGGCTCTGTATTCACAGCCATGGAAT	1457		
Db	541 ACTCAGTGTATCTGGAAGACCGCGCCGCCAAAGAGGCTCTGTATTCACAGCCATGGAAT	600		
1458	CCATCCAGGAGAGGCTGCAGCGGGGTGGTACCATGAGAGCTATGCAATATGGGGGGA	1517		
Db	601 CCATCCAGGAGAGGCTGCAGCGGGGTGGTACCATGAGAGCTATGCAATATGGGGGGA	660		
1518	TCCCTGCGCCGGAATAGTAAAGGGAAGGCTTCTGCTTATATTTGGCATCATTTGACATTC	1577		
Db	661 TCCCTGCGCCGGAATAGTAAAGGGAAGGCTTCTGCTTATATTTGGCATCATTTGACATTC	720		
1578	TACACTCTTACAGG-TTTGGTTAAGAAATTGGAGCACTCTTGGAAAGCCCTGGTACATGAC	1636		
Db	721 TACACTCTTACAGG-TTTGGTTAAGAAATTGGAGCACTCTTGGAAAGCCCTGGTACATGAC	780		

QY	1637	GGAGGACACTGCTCAGTGCATGCCAGGCTCTACGCTGAAGCGTTCAGGCTTCATG	1696
Db	781	GGAGGACACTGCTCAGTGCATGCCAGGCTCTACGCTGAAGCGTTCAGGCTTCATG	840
QY	1697	TCGCAACACAGTATTTAAGAAGATTCCTTGAAGCCTTCCTCCATCAAAA	1745
Db	841	TCGCCACACAGTATTTAAGAAGATTCCTTGAAGCCTTCCTCCATCAAAA	889
RESULT 2			
LOCUS	BM562816	1045 bp	mRNA
DEFINITION	AGENCOURT_6588557 NIH_MGC_98 Homo sapiens	cdna clone IMAGE:5478529	
ACCESSION	BM562816		
VERSION	BM562816.1	GI:18809218	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 1045)		
TITLE	NIH-MGC <a href="http://nmc.ncbi.nih.gov/">http://nmc.ncbi.nih.gov/</a> .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: L10CMI98 row: f column: 02 High quality sequence stop: 634. Location/Qualifiers		
FEATURES	1..1045		
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="5478529" /clone_11d="NIH_MGC_98" /tissue_type="astrocytoma grade IV, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: brain; Vector: pOTB1; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."		
BASE COUNT	259 a 270 c 238 g 278 t		
ORIGIN			
Query Match	22.7%;	Score 84.1;	DB 13;
Best Local Similarity	93.2%;	Pred. No. 5.2e-183;	
Matches 914;	Conservative 0;	Mismatches 60;	Indels 7;
		Gaps 3	
QY	1445	ACAGCAGTGAATTCATCCAGGAGAGAGGCTGCACGCGGCTGTACCATGAGAGCTGATGAC	1504
Db	1	ACAGCAGTGAATTCATCCAGGAGAGGCTGCACGCGGCTGTACCATGAGAGCTGATGAC	60
QY	1505	CATATGGGTGGCATCCCTGCGCGGAATAGTAAGGGGAAAGGCTTCTGCTTATATATGGC	1564
Db	61	CATATGGGTGGCATCCCTGCGCGGAATAGTAAGGGGAAAGGCTTCTGCTTATATATGGC	120
QY	1565	ATCATTTGACATTTCTACAGCTTACAGGTTGTTAAGAAAGTTGAGACACTTGGAAAGCC	1624
Db	121	ATCATTTGACATTTCTACAGCTTACAGGTTGTTAAGAAAGTTGAGACACTTGGAAAGCC	180
QY	1625	CTGTGACATGACGAGACACTGTCTCAGTGCATGCCAGGCTTCATACCTTAACGGTTC	1684

QY	1637	GGAGGACACTGCTCAGTGCATGCCAGGCTCTACGCTGAAGCGTTCAGGCTTCATG	1696
Db	781	GGAGGACACTGCTCAGTGCATGCCAGGCTCTACGCTGAAGCGTTCAGGCTTCATG	840
QY	1697	TCGCAACACAGTATTTAAGAAGATTCCTTGAAGCCTTCCTCCATCAAAA	1745
Db	841	TCGCCACACAGTATTTAAGAAGATTCCTTGAAGCCTTCCTCCATCAAAA	889
RESULT 2			
LOCUS	BM562816	1045 bp	mRNA
DEFINITION	AGENCOURT_6588557 NIH_MGC_98 Homo sapiens	cdna clone IMAGE:5478529	
ACCESSION	BM562816		
VERSION	BM562816.1	GI:18809218	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 1045)		
TITLE	NIH-MGC <a href="http://nmc.ncbi.nih.gov/">http://nmc.ncbi.nih.gov/</a> .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: L10CMI98 row: f column: 02 High quality sequence stop: 634. Location/Qualifiers		
FEATURES	1..1045		
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="5478529" /clone_11d="NIH_MGC_98" /tissue_type="astrocytoma grade IV, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: brain; Vector: pOTB1; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."		
BASE COUNT	259 a 270 c 238 g 278 t		
ORIGIN			
Query Match	22.7%;	Score 84.1;	DB 13;
Best Local Similarity	93.2%;	Pred. No. 5.2e-183;	
Matches 914;	Conservative 0;	Mismatches 60;	Indels 7;
		Gaps 3	
QY	1445	ACAGCAGTGAATTCATCCAGGAGAGAGGCTGCACGCGGCTGTACCATGAGAGCTGATGAC	1504
Db	1	ACAGCAGTGAATTCATCCAGGAGAGGCTGCACGCGGCTGTACCATGAGAGCTGATGAC	60
QY	1505	CATATGGGTGGCATCCCTGCGCGGAATAGTAAGGGGAAAGGCTTCTGCTTATATATGGC	1564
Db	61	CATATGGGTGGCATCCCTGCGCGGAATAGTAAGGGGAAAGGCTTCTGCTTATATATGGC	120
QY	1565	ATCATTTGACATTTCTACAGCTTACAGGTTGTTAAGAAAGTTGAGACACTTGGAAAGCC	1624
Db	121	ATCATTTGACATTTCTACAGCTTACAGGTTGTTAAGAAAGTTGAGACACTTGGAAAGCC	180
QY	1625	CTGTGACATGACGAGACACTGTCTCAGTGCATGCCAGGCTTCACCTTACAGGTTTC	1684

Db 181 CTGCTACATGACGAGACACTGCTCATGATCCGCCAGGCTTCTACGCTGAAGGTTTC 240  
 Oy 1685 CAGGGCTTCATGTCACACAGATATTTAGAAATTCCTTGAACCTTCCCTCCAA 1744  
 Db 241 CAGGCTTCATGTCACACAGATATTTAGAAATTCCTTGAACCTTCCCTCCAA 300  
 Oy 1745 AAGTTTGGTGGCTCATCTTCTCTGGGAGAGAGGCTCCAGTGGCACTCTGCATT 1804  
 Db 301 AAGTTTGGTGGCTCATCTTCTCTGGGAGAGAGGCTCCAGTGGCACTCTGCATT 360  
 Oy 1805 ACTTACACGACCATGCTCTGCGGACACAAAGGACAGTACACAAAGGACAGAGTG 1864  
 Db 361 ACTTACACGACCATGCTCTGCGGACACAAAGGACAGTACACAAAGGACAGAGTG 420  
 Oy 1865 GAGGACGAGCTTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1924  
 Db 421 GAGGACGAGCTTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 480  
 Oy 1925 ATCAGTGGAGGCTGGCTTATTCCTGACCCAGTTCCTGACCTGATGGAGAGACTTGG 1984  
 Db 481 ATCAGTGGAGGCTGGCTTATTCCTGACCCAGTTCCTGACCTGATGGAGAGACTTGG 540  
 Oy 1985 CAAATGCTAATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2044  
 Db 541 CAAATGCTAATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 600  
 Oy 2045 CAAATGCTAATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2104  
 Db 601 CAAATGCTAATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 660  
 Oy 2105 ATGTCAGGCTTGGCCCGACGATGCTGAATTTCTTCTACTTGGTACATCAAAAAGAGT 2164  
 Db 661 ATGTCAGGCTTGGCCCGACGATGCTGAATTTCTTCTACTTGGTACATCAAAAAGAGT 720  
 Oy 2165 GTAAATGAGTGGAGGAGGCTGCTCTCCATCTTCTCTGGAAGAAGCTTCTCTCT 2224  
 Db 721 GTAAATGAGTGGAGGAGGCTGCTCTCCATCTTCTCTGGAAGAAGCTTCTCTCT 780  
 Oy 2225 TCCCTTCCCTCATGATGATGGGCTTATGAGTGGTGGTGGTGGTGGTGGTGGTGGT 2284  
 Db 781 TCCCTTCCCTCATGATGATGGGCTTATGAGTGGTGGTGGTGGTGGTGGTGGTGGT 840  
 Oy 2285 CACTCCAGAGTGGGCTGG--TACGGAATTTTCAACTGGCC---AACCTTGGCTCCACT 2338  
 Db 841 CACTCCAGAGTGGGCTGGTGGGATTTTAAAGGGCCCAACCCCTTGGCTCCACT 900  
 Oy 2339 ATT-GAATTTTTCACAGCCCATTTCTTCATGCTGGAATGGAGTGGCTGGACTTGGCA 2397  
 Db 901 ATTGGAATTTTTCACAAACCCCTTCTTCATGATGATGGAATAATGGGAATTTGGCT 960  
 Oy 2398 GCTTCTTTCCTCCCTGCTTT 2418  
 Db 961 GGGGACTTGGGAAGGTTT 981

RESULT 3  
 BM545034 1119 bp mRNA linear EST 20-FEB-2002  
 LOCUS ABENCOUR\_6497385 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:558832  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM545034  
 VERSION BM545034.1 GI:18776787  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1119)  
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Invitrogen  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ULNL at:  
<http://image.llnl.gov>  
 Plate: LAM12360 row: j column: 01  
 High quality sequence stop: 32.  
 High quality sequence stop: 652.  
 location/Qualifiers  
 1. 1119  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:558832"  
 /clone\_11b="NIH\_MGC\_125"  
 /lab\_host="DH10B"  
 /note="Organ: ovary (pool of 3); Vector: pCMV-Sport6;  
 site\_1: EcoRV (destroyed); site\_2: NotI; RNA source pool  
 of three ovaries, from females ranging in age from 38 to  
 49 yo. Library is oligo-dT primed and directionally cloned  
 (EcoRV site is destroyed upon cloning). Average insert  
 size 2.1 kb, insert size range 1-3.5 kb. Library is  
 normalized and enriched for full-length clones and was  
 constructed by C. Gruber (Invitrogen). Research Genetics  
 tracking code 036."

BASE COUNT 276 a 239 c 246 g 357 t 1 others  
 ORIGIN

Query Match 22.3%; Score 828.2; DB 13; Length 1119;  
 Best Local Similarity 94.8%; Pred. NO. 4.7e-180;  
 Matches 890; Conservative 0; Mismatches 43; Indels 6; Gaps 3;

Oy 2780 CTGAAGGAATGAGAGAGTGGAGCATGGGGTAACTTTATCCCTTTGTTAAACAGAG 2839  
 Db 16 CGATGGAATGAGAGAGTGGAGCATGGGGTAACTTTATCCCTTTGTTAAACAGAG 75  
 Oy 2840 GCAGCATGGGCTGGAGATCATAGCCCTTCTAGGCAAAATCCTGACAGGCT 2899  
 Db 76 GCAGCATGGGCTGGAGATCATAGCCCTTCTAGGCAAAATCCTGACAGGCT 135  
 Oy 2900 ATAGTAATTTACTATTTTGAATTAATATTTGTTGTTTCTAAATGTGA 2959  
 Db 136 ATAGTAATTTACTATTTTGAATTAATATTTGTTGTTTCTAAATGTGA 195  
 Oy 2960 GACTTACCAATGAATTTAGATCATCTCCAGAGAGATTTTCTGCTTTCATCT 3019  
 Db 196 GACTTACCAATGAATTTAGATCATCTCCAGAGAGATTTTCTGCTTTCATCT 255  
 Oy 3020 TTTCACAGTGTCTCCCTGTTTGGAGCTAAGGTAAGAGGGGCACTTGTGCTGT 3079  
 Db 256 TTTCACAGTGTCTCCCTGTTTGGAGCTAAGGTAAGAGGGGCACTTGTGCTGT 315  
 Oy 3080 TTAACAGAGTCCATATCTGTGAGCCAGCAAAATTTTCTTAACTCATGGGAGACAG 3139  
 Db 316 TTAACAGAGTCCATATCTGTGAGCCAGCAAAATTTTCTTAACTCATGGGAGACAG 375  
 Oy 3140 CAGATTTCTGCTTGTGAGTGCATCTGCTGCCATATGCTTACCCCTGCTTCATG 3199  
 Db 376 CAGATTTCTGCTTGTGAGTGCATCTGCTGCCATATGCTTACCCCTGCTTCATG 435  
 Oy 3200 CAGGGAAGTGGAAATGGGGGCTACATATAGCCCTCCCGGCTCAAAAGTTGTG 3259  
 Db 436 CAGGGAAGTGGAAATGGGGGCTACATATAGCCCTCCCGGCTCAAAAGTTGTG 495  
 Oy 3260 TTTTCATCTGATCCCTTCCACTTGTGAGGGGAAGAGGGGCTGATCTCAGGACAG 3319  
 Db 496 TTTTCATCTGATCCCTTCCACTTGTGAGGGGAAGAGGGGCTGATCTCAGGACAG 555  
 Oy 3320 ATTGTTGAATTTCTGTTTATCTCTTATCCCAACCTGCTTGATTAATATGTTAGCC 3379  
 Db 556 ATTGTTGAATTTCTGTTTATCTCTTATCCCAACCTGCTTGATTAATATGTTAGCC 615

QY 3380 CATACCCCAATTAAGTCTATATTAAGACACACCCAGCTTTCGCTGCTCTTT 3439  
 |||||||  
 Db 616 CATACCCCAATTAAGTCTATATTAAGACACCCAGCTTTCGCTGCTCTTT 675  
 QY 3440 GCTGCCATGTTTTTACAGAGAAAGAAATCTGCTATTTTTTTCATTAATTA 3499  
 |||||||  
 Db 676 GCTGCCATGTTTTTACAGAGAAAGAAATCTGCTATTTTTTTCATTAATTA 735  
 QY 3500 TTTATGATGATTAAGTCTTATTAAGACAGAGTTCGTAGGGTGGGAGATA 3559  
 |||||||  
 Db 736 TTTATGATGATTAAGTCTTATTAAGACAGAGTTCGTAGGGTGGGAGATA 795  
 QY 3560 TTT-----GAGGAGAGGCTGGCTCTTACGGAAGAAATGGGGAACATTTTTTAA 3615  
 |||||||  
 Db 796 TTTTGAAGGAGGGGCTGGCTCTTACGGAAGAAATGGGGAACATTTTTTAA 855  
 QY 3616 TGTACTATTTTG-CCTTACTTTTGTATTTTCAG-AAATGGCAATTAATTAAGTG 3673  
 |||||||  
 Db 856 GGGGACTATTTTGGCCCTCCACTTTGCAATTTGCAAGAAATGGCAATTAATTAAG 915  
 QY 3674 ATATAGGTTTTTAATGATTAATTAAGTATTT 3712  
 |||||||  
 Db 916 TTTATGCTTTTATTAATTAAGTATTAAGTATTT 954

RESULT 4  
 AL522833/c 865 bp mRNA linear EST 13-FEB-2001  
 LOCUS AL522833 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CSDB009YB17 3  
 DEFINITION AL522833 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CSDB009YB17 3  
 ACCESSION AL522833  
 VERSION AL522833.1 GI:12786326  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 865)  
 AUTHORS Li, W. B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 101 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 FEATURES  
 source  
 1. 865  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSDB009YB17"  
 /clone\_1bp="LTI\_NFL004\_NBC2"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng liang life technologies, a division of invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 280 a 185 c 180 g 213 t 7 others  
 ORIGIN

Query Match 22.3%; Score 827; DB 9; Length 865;  
 Best Local Similarity 97.6%; Pred. No. 8.6e-180;  
 Matches 844; Conservative 4; Mismatches 16; Indels 1; Gaps 1;  
 QY 2809 GTAATCTTAATCCCTTTGTTAAACAGCA-GGACAGCATGGCTGGAGATCATATAGCC 2867

Db 865 GTAATCTTAATCCCTTTGTTAAACAGCA-GGACAGCATGGCTGGAGATCATATAGCC 806  
 QY 2868 TTCCATGGAGAAATCCCTGTTCACTGACAGGCTATTAATTAATTAATTTGCAATTTG 2927  
 |||||||  
 Db 805 TTCCATGGAGAAATCCCTGTTCACTGACAGGCTATTAATTAATTAATTTGCAATTTG 746  
 QY 2928 AAATATATCTGCTGTTGTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 2987  
 |||||||  
 Db 745 AAATATATCTGCTGTTGTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 686  
 QY 2988 TCCAGAGAGATTTTTTTTGGCTCTTCTCATCTTTTCCACAGTGTCTGCTTTTGA 3047  
 |||||||  
 Db 685 TCCAGAGAGATTTTTTTTGGCTCTTCTCATCTTTTCCACAGTGTCTGCTTTTGA 626  
 QY 3048 GCTAAGTAAAGAGGAGACCTCTGCTGTTTAAACAGACAGTATCTGTGAGGCA 3107  
 |||||||  
 Db 625 GCTAAGTAAAGAGGAGACCTCTGCTGTTTAAACAGACAGTATCTGTGAGGCA 566  
 QY 3108 GCAAAATATTTTCTTAACATCATGAGGAGACAGATTTCTGCTGTGAGGCTATTC 3167  
 |||||||  
 Db 565 GCAAAATATTTTCTTAACATCATGAGGAGACAGATTTCTGCTGTGAGGCTATTC 506  
 QY 3168 TGTGCCATATGCTCTACCCCTGCTCTCATGAGGAGAAATGGGCTACATA 3227  
 |||||||  
 Db 505 TGTGCCATATGCTCTACCCCTGCTCTCATGAGGAGAAATGGGCTACATA 446  
 QY 3228 TGCCCTCTCTCCCGCTCAAGAGTGTGTTTTCATCTGATCTTCACTCTGTC 3287  
 |||||||  
 Db 445 TGCCCTCTCTCCCGCTCAAGAGTGTGTTTTCATCTGATCTTCACTCTGTC 386  
 QY 3288 AGGGGAAAGAGGGGCTGATCTCAGCAGATTTGTAATTCCTGTTCTATCCCTCT 3347  
 |||||||  
 Db 385 AGGGGAAAGAGGGGCTGATCTCAGCAGATTTGTAATTCCTGTTCTATCCCTCT 326  
 QY 3348 CTATCCACCTGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3407  
 |||||||  
 Db 325 CTATCCACCTGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 266  
 QY 3408 CACCCCGAGCAGTTTCTGCTGCTGCTCTTGTGCTGCAATTTTTCACAGAGAAAG 3467  
 |||||||  
 Db 265 CMCCCGAGCAGTTTCTGCTGCTGCTCTTGTGCTGCAATTTTTCACAGAGAAAG 206  
 QY 3468 AATTTCTGCTATTTTCTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTA 3527  
 |||||||  
 Db 205 AATTTCTGCTATTTTCTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTA 146  
 QY 3528 GGACAGAGTCTGTTAGGGGAGGAGGAAATTTTGGAGGAGGCTGCTTACGGAA 3587  
 |||||||  
 Db 145 GGACAGAGTCTGTTAGGGGAGGAGGAAATTTTGGAGGAGGCTGCTTACGGAA 86  
 QY 3588 GGAATGGGAGAGCAATTTTATTAAGTCTTACATTTGCTCTACTTTGTAATTTCA 3647  
 |||||||  
 Db 85 GGAATGGGAGAGCAATTTTATTAAGTCTTACATTTGCTCTACTTTGTAATTTCA 26  
 QY 3648 GAAATGGCAATTAATTAATTAAGT 3672  
 |||||||  
 Db 25 GAAATGGCAATTAATTAATTAAGT 1  
 |||||||  
 RESULT 5  
 AUI32382 838 bp mRNA linear EST 01-NOV-2002  
 LOCUS AUI32382 NT2RP3 Homo sapiens cDNA clone NT2RP3004316 5', mRNA  
 DEFINITION AUI32382 NT2RP3 Homo sapiens cDNA clone NT2RP3004316 5', mRNA  
 ACCESSION AUI32382  
 VERSION AUI32382.1 GI:10992736  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 838)

**AUTHORS** Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.

**TITLE** HRI human cDNA project

**JOURNAL** Unpublished (2000)

**COMMENT** Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp

**FEATURES** HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

**location/Qualifiers**

**source**

1. 838

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="NT2RP3004316"

/clone\_11b="NT2RP3"

/cell\_type="teratocarcinoma"

/note="Vector: pME18SF13; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"

**BASE COUNT** 227 a 200 c 199 g 210 t 2 others

**ORIGIN**

Query Match 22.1%; Score 820.2; DB 9; Length 838;  
Best Local Similarity 99.5%; Pred. No. 3.2e-178;  
Matches 833; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Oy 827 ATCCGGCCCGAGTACTTGTATCCCTGCGAGTGAAGCCGCTGATGACTGTAGC 886  
|||||  
Db 1 ATCCGGCCCGAGTACTTGTATCCCTGCGAGTGAAGCCGCTGATGACTGTAGC 60  
Oy 887 TCTGAGCTAGTGTGTCCTATTCATTCATGTCACGACGATGATGATTAAGACA 946  
|||||  
Db 61 TCTGAGCTAGTGTGTCCTATTCATTCATGTCACGACGATGATGATTAAGACA 120  
Oy 947 GTCCAACTTAAGAGCGGAGTTCGACGAAGCTGCTCCAGATCTACATGAACCTC 1006  
|||||  
Db 121 GTCCAACTTAAGAGCGGAGTTCGACGAAGCTGCTCCAGATCTACATGAACCTC 180  
Oy 1007 AACCGAACCTCGGACTTGGCTTAAATCTATGACGACTGCTGCGAGCAGGT 1066  
|||||  
Db 181 AACCGAACCTCGGACTTGGCTTAAATCTATGACGACTGCTGCGAGCAGGT 240  
Oy 1067 GGCAAGAACATTCGATTTGTGTGATGAACAATCTTTTACCAAGATCGGTAAAAATGCAT 1126  
|||||  
Db 241 GGCAAGAACATTCGATTTGTGTGATGAACAATCTTTTACCAAGATCGGTAAAAATGCAT 300  
Oy 1127 ATCAAAATGACCTTAAGAGCTCAACCTACAAAGCGCGGCTTCCAGAAAGAGCGAGAG 1186  
|||||  
Db 301 ATCAAAATGACCTTAAGAGCTCAACCTACAAAGCGCGGCTTCCAGAAAGAGCGAGAG 360  
Oy 1187 AAGCCTCTCCACATTTAAAGACCTAGACTCTTCAAGACATCCGATGGTCTTTT 1246  
|||||  
Db 361 AAGCCTCTCCACATTTAAAGACCTAGACTCTTCAAGACATCCGATGGTCTTTT 420  
Oy 1247 TTGATGCTGACATGTACAGCCTCTCTGTAGAGACCTGACGCTGATGTTGGTGGC 1306  
|||||  
Db 421 TTGATGCTGACATGTACAGCCTCTCTGTAGAGACCTGACGCTGATGTTGGTGGC 480  
Oy 1307 CAGAGCTCAAGATTAATGATTTACAGCCTCTGATGTCATTCATTAATATGATATGCA 1366  
|||||  
Db 481 CAGAGCTCAAGATTAATGATTTACAGCCTCTGATGTCATTCATTAATATGATATGCA 540  
Oy 1367 CAACGAGAGCCCTTAAGCAGTGAACACAGTACTAGTTGATGATCGAAGAGCGGCCCC 1426  
|||||  
Db 541 CAACGAGAGCCCTTAAGCAGTGAACACAGTACTAGTTGATGATCGAAGAGCGGCCCC 600

Oy 1427 CAAAAGCTCTGTATTCACAGCCATGAAATCCATCCAGGAGAGGCTCGACGGGTGCT 1486  
|||||  
Db 601 CAAAAGCTCTGTATTCACAGCCATGAAATCCATCCAGGAGAGGCTCGACGGGTGCT 660  
Oy 1487 ACCATGAGACATGATGACATATGATGGTGCATCCCTGCCCGAAATGATAAGGGAAGG 1546  
|||||  
Db 661 ACCATGAGACATGATGACATATGATGGTGCATCCCTGCCCGAAATGATAAGGGAAGG 720  
Oy 1547 CTTTCGCTTTATATTTGGCATTCATGACATCTACAGTCTTACAGGTTTGTAAAGATTG 1606  
|||||  
Db 721 CTTTCGCTTTATATTTGGCATTCATGACATCTACAGTCTTACAGGTTTGTAAAGATTG 780  
Oy 1607 GAGCACTCTTGAAGAACCCCTGCTACATGACGAGACACTGTCTCAGTGCATGCCCA 1663  
|||||  
Db 781 GAGCACTCTTGAAGAACCCCTGCTACATGACGAGAC -CTGGCTCAGTGCATGCCCA 836

**RESULT** 6

**BM549810** 1069 bp mRNA linear EST 20-PEB-2002

**LOCUS** AGENCOURT.6543938 NIH\_MGC\_118 Homo sapiens cDNA clone IMAGE:5745824

**DEFINITION** 5', mRNA sequence.

**ACCESSION** BM549810

**VERSION** BM549810.1 GI:18785525

**KEYWORDS** EST.

**SOURCE** human.

**ORGANISM** Homo sapiens

**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** 1 (bases 1 to 1069)

**TITLE** NIH-MGC http://mhc.nci.nih.gov/.

**JOURNAL** National Institutes of Health, Mammalian Gene Collection (MGC)

**COMMENT** Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM12769 row: 9 column: 09  
High quality sequence start: 19  
High quality sequence stop: 574.

**FEATURES** location/Qualifiers

**source**

1. 1069

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5745824"

/clone\_11b="NIH\_MGC\_118"

/tissue\_type="leukocyte"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH\_MGC Library."

**BASE COUNT** 263 a 290 c 216 g 298 t 2 others

**ORIGIN**

Query Match 22.0%; Score 817; DB 13; Length 1069;  
Best Local Similarity 96.8%; Pred. No. 1.8e-177;  
Matches 865; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

Oy 1725 TGAAGCTTCTCCCTTCCAAAAGTTTCGCTGCTCATCTTCTCCGCGACGAGCT 1784  
|||||  
Db 35 TGAAGCTTCTCCCTTCCAAAAGTTTCGCTGCTCATCTTCTCCGCGACGAGCT 93

QY 1785 CCAGTGGCAACTCTGCTGATTACTTACCAAGCCATCGTCTCTGGGGAACACAAAGCACAG 1844  
 DB 94 CCAAGTGGCAACTCTGCTGATTACTTACCAAGCCATCGTCTCTGGGGAACACAAAGCACAG 153  
 QY 1845 TGACAAACAAAGGAGAAAGTGGAGCCGCTTACCTTGGTCTGCTGATGTTTACCTC 1904  
 DB 154 TGACAAACAAAGGAGAAAGTGGAGCCGCTTACCTTGGTCTGCTGATGTTTACCTC 213  
 QY 1905 AGAGCTCCAGCTTGGAGGAATCAATGAGGAGCGCTTATCTGAGCCAGTTTCTCAC 1964  
 DB 214 AGAGCTCCAGCTTGGAGGAATCAATGAGGAGCGCTTATCTGAGCCAGTTTCTCAC 273  
 QY 1965 CTCTAGTTGGAGAGACTTTCGAAATGCTAACTACAGTACACCTTGGAAAGCTTGAAG 2024  
 DB 274 CTCTAGTTGGAGAGACTTTCGAAATGCTAACTACAGTACACCTTGGAAAGCTTGAAG 333  
 QY 2025 TTGAGAGTCAAGATTAACCATTAAGCCGAAAGCCTCAGAAAGCTTGGAAAGATTC 2084  
 DB 334 TTGAGAGTCAAGATTAACCATTAAGCCGAAAGCCTCAGAAAGCTTGGAAAGATTC 393  
 QY 2085 GGCATCTCTGATCCCAAGATGTCAGCCCTTGGCCGAGAAATGTAATTTTCTCTAC 2144  
 DB 394 GGCATCTCTGATCCCAAGATGTCAGCCCTTGGCCGAGAAATGTAATTTTCTCTAC 453  
 QY 2145 TTGCTCATCAAAAAAGAGTGTAAATAGAGTGAAGGAGCTGCTCTCATCTTCTTCT 2204  
 DB 454 TTGCTCATCAAAAAAGAGTGTAAATAGAGTGAAGGAGCTGCTCTCATCTTCTTCT 513  
 QY 2205 GAAGAAGAACCTTCT 2264  
 DB 514 GAAGAAGAACCTTCT 573  
 QY 2265 GAGAGCCGACATCCCTCCACCTCCAGAGTGGGTGGTACGATTTTCAACTGGCCAAC 2324  
 DB 574 GAGAGCCGACATCCCTCCACCTCCAGAGTGGGTGGTACGATTTTCAACTGGCCAAC 633  
 QY 2325 CCTTGGCTCCACATTTGAATTTTTCAGACCCCAATCTTCAATGCTGGAATGGGAT 2384  
 DB 634 CCTTGGCTCCACATTTGAATTTTTCAGACCCCAATCTTCAATGCTGGAATGGGAT 693  
 QY 2385 GCTGAGCTTGGAGAGCTTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2444  
 DB 694 GCTGAGCTTGGAGAGCTTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 753  
 QY 2445 TCAGACAGACTAGCTGGACATTAATCCCTTACCTTCTCTCTCTCTCTCTCTCTCTCT 2504  
 DB 754 TCAGACAGACTAGCTGGACATTAATCCCTTACCTTCTCTCTCTCTCTCTCTCTCTCT 813  
 QY 2505 GAATACCTCTGTAAATCTGTAAAGG-ATTTTGGGGGATTAAGGGTTTAAACCACTCC 2563  
 DB 814 -AATACCTCTGTAAATCTGTAAAGGTTTGGGGGATTAAGGGTTTAAACCACTCC 872  
 QY 2564 AGCTTCT 2617  
 DB 873 AACTTCT 926  
 RESULT 7 824 bp mRNA linear EST 01-AUG-2002  
 LOCUS AUI26052 NT2RM4 Homo sapiens cDNA clone NT2RM4002631 5', mRNA  
 DEFINITION sequence.  
 ACCESSION AUI26052.1 GI:10950768  
 VERSION AUI26052.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 824)  
 AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Isogai,T.

TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,  
 Y., Sugano,S., Isogai,T.)  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: genomics@hri.co.jp  
 HRI human cDNA project: 5'-3'-end one pass sequencing: Helix  
 Research Institute; cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 FEATURES  
 source 1..824  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2RM4002631"  
 /clone\_1lb="NT2RM4"  
 /cell\_type="teratocarcinoma"  
 /cell\_line="NT2"  
 /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
 precursor cells"  
 BASE COUNT 223 a 201 c 189 g 207 t 4 others  
 ORIGIN  
 Query Match 21.9%; Score 813.6; DB 9; Length 824;  
 Best local Similarity 99.0%; Pred. No. 1.1e-176;  
 Matches 816; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 721 CTTTCCAGTGAAGGAGGACCACTGACCCCTGCTCATCTACATGACTTTCGTTCAA 780  
 DB 1 CTTTCCAGTGAAGGAGGAGCAACTGACCCCTGCTCATCTACATGACTTTCGTTCAA 60  
 QY 781 GACCTATGCACTGTTGCTTCCGCTACTTCCGGAGCTATTGTGATCCGGCCGATGA 840  
 DB 61 GACCTATGCACTGTTGCTTCCGCTACTTCCGGAGCTATTGTGATCCGGCCGATGA 120  
 QY 841 TTACTGTATTCCTCTGACAGTACCGCGATTAATCTGTAGCTCTGAGCTAGTGG 900  
 DB 121 TTACTGTATTCCTCTGACAGTACCGCGATTAATCTGTAGCTCTGAGCTAGTGG 180  
 QY 901 TTCCCTATTTATGTGTCACGACGATGATCATTTAAGACAGTCCACATTAAGA 960  
 DB 181 TTCCCTATTTATGTGTCACGACGATGATCATTTAAGACAGTCCACATTAAGA 240  
 QY 961 GGGGAAATTCGAGAGAGTGTCCAGAGATCTACATGACCTAACCAAGAACCTCG 1020  
 DB 241 GGGGAAATTCGAGAGAGTGTCTCCAGAGATCTACATGACCTAACCAAGAACCTCG 300  
 QY 1021 GACTTCTGCTTAATTTCTATGAGACTGTACTGTGTGACGAGGAGGCAAGCAATTGG 1080  
 DB 301 GACTTCTGCTTAATTTCTATGAGACTGTACTGTGTGACGAGGAGGCAAGCAATTGG 360  
 QY 1081 GATTGTGATGTAAGCAATCTTTTACCAAGTGGTAAATGCAATTCATATGAGCT 1140  
 DB 361 GATTGTGATGTAAGCAATCTTTTACCAAGTGGTAAATGCAATTCATATGAGCT 420  
 QY 1141 CAAAGGCTCAACTTCAAGAGGCGGCTTCCAGAAAGACGAGAGAGCTTCTCCAC 1200  
 DB 421 CAAAGGCTCAACTTCAAGAGGCGGCTTCCAGAAAGACGAGAGAGCTTCTCCAC 480  
 QY 1201 ATTTAAGACTTAAGCTTTTCAAGAGATCCCTGATGCTTTTGGATGCTGACAT 1260  
 DB 481 ATTTAAGACTTAAGCTTTTCAAGAGATCCCTGATGCTTTTGGATGCTGACAT 540  
 QY 1261 GTACAGAGCTCTGTAAAGCTGACAGCTGATGTTTGGTGTGAGAGCTTCAAGAT 1320  
 DB 541 GTACAGAGCTCTGTAAAGCTGACAGCTGATGTTTGGTGTGAGAGCTTCAAGAT 600  
 QY 1321 AATGATTACAGCTCTTGTATGCAATCCATTAATATGATCATGACACAGAGAGCCCTT 1380



QY	2974	ATTTTAGATCATTTCTCCAGAGGAGATTTTTTTTGGCTCTCTCTCATCTTTTCCAAACAGTGT	3033
Db	634	ATTTTAGATCATTTCTCCAGAGGAGATTTTTTTTGGCTCTCTCTCATCTTTTCCAAACAGTGT	575
QY	3034	CTCTGTGTTCTGGAGCTTAAGGTAACAGAGGGGACACITCTGTCTGTGTTTAAACACACAGTCCA	3093
Db	574	CTCTGTGTTCTGGAGCTTAAGGTAACAGAGGGGACACITCTGTCTGTGTTTAAACACACAGTCCA	515
QY	3094	TATCTGTGAGGCCAGCAAAATATTTTCTTAAACTCATGGGGAGACAGAGATTTCTTGCCTT	3153
Db	514	TATCTGTGAGGCCAGCAAAATATTTTCTTAAACTCATGGGGAGACAGAGATTTCTTGCCTT	455
QY	3154	GGTGAGGTCAATGCTGTGCCATATGTCCTACCCCTGTCTTCATGCGAGGGAAGTTGGAA	3213
Db	454	GGTGAGGTCAATGCTGTGCCATATGTCCTACCCCTGTCTTCATGCGAGGGAAGTTGGAA	395
QY	3214	ATGGGGCTACATATGCCCCCTCTCTCCCGCTCTACAAGAGTTGTGGTTTTCCTCATCTGATC	3273
Db	394	ATGGGGCTACATATGCCCCCTCTCTCCCGCTCTACAAGAGTTGTGGTTTTCCTCATCTGATC	335
QY	3274	CTTCCACTCTGTGAGGGGAAGAGGGGCGCTGGTATCTCAGGCAGATTTGTTGAATCCT	3333
Db	334	CTTCCACTCTGTGAGGGGAAGAGGGGCGCTGGTATCTCAGGCAGATTTGTTGAATCCT	275
QY	3334	GTTCTATCCCTTCTCTATCCACCCCTGCCTTGATATATGTTAGCCCATACCCCAATAA	3393
Db	274	GATCTAMCCCBTCTCTATCCACCCCTGCCTAGATAATATGTTAGCCCATACCCCAATAA	215
QY	3394	CTGTCTATATAGACACCCCGCCAGCTTTCTGGCTGCCTGTCTTGTGCGCATCTTTTT	3453
Db	214	CTGTCTATATAGACACCCCGCCAGCTTTCTGGCTGCCTGTCTTGTGCGCATCTTTTT	155
QY	3454	TACAAGAAGAAAGAAATCTGTCTATTTTTTTTTTTCATAATTTACTATTTATGATGTATTT	3513
Db	154	TACAAGAAGAAAGAAATCTGTCTATTTTTTTTTTTCATAATTTACTATTTATGATGTATTT	95
QY	3514	AAGTGTTTTTATTAAGGACAGATTCGTGTAGGGGTGGGAGGAATATTTTGAGGAGGCGT	3573
Db	94	AAGTGTTTTTATTAAGGACAGATTCGTGTAGGGGTGGGAGGAATATTTTGAGGAGGCGT	35
QY	3574	GGGTCTTAGGAAAGGAATGGGGAAGCAACATTT 3607	
Db	34	GGGCTTTAGGAAAGGAATGGGGAAGCAACATTT 1	
RESULT 9			
LOCUS	BQ439604	904 bp mRNA linear EST 24-MAY	
DEFINITION	AGENCOURT_7828840 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6055		
ACCESSION	BQ439604		
VERSION	BQ439604.1	GI:21178680	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 904)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-r@mail.nih.gov		
	Tissue procurement: DCTD/Drp/Gazdar		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://Image.llnl.gov		
	Plate: LLAM13318 row: d column: 11		
	High quality sequence stop: 693.		

[illegible]

241 AGGAGCGGCTCGACGTCTGTAGGAGG - CCCGAGGGGCGGGAGGTGCGCCACAGA 299  
 280 AGGAGCGGCTCGACGTCTGTAGGAGGCGCCGAGGGGCGGGAGGTGCGCCACAGA 339  
 300 ACGGGGTTCTGTAAGAGAGCGTTGGGAAGATTGATTCGAGAGAGAGAGAAACCGGAT 359  
 340 ACGGGGTTCTGTAAGAGAGCGTTGGGAAGATTGATTCGAGAGAGAGAGAAACCGGAT 399  
 360 TGAAGAGAGCGGCGCTGAGGGGAGGGGCTGCTAAGATGGCGTGGCGCTCTCCG 419  
 400 TGAAGAGAGCGGCGCTGAGGGGAGGGGCTGCTAAGATGGCGTGGCGCTCTCCG 459  
 420 GGCGTCTGCTTCGGTGGTTTTCATCCTTTGATCCGCGGCTCCCTTCCTGCTTGT 479  
 460 GGCGTCTGCTTCGGTGGTTTTCATCCTTTGATCCGCGGCTCCCTTCCTGCTTGT 519  
 480 CCTCAGCATCTGGAATCAAGAGAGCCATGGCATCTGAGGTGCTTATGCCCTCTGGCATGC 539  
 520 CCTCAGCATCTGGAATCAAGAGAGCCATGGCATCTGAGGTGCTTATGCCCTCTGGCATGC 579  
 540 CCATCAAGAAAATAGGCATAGAAGTGTGATTCCTCAGGAGAGACACATATAAAGA 599  
 580 CCATCAAGAAAATAGGCATAGAAGTGTGATTCCTCAGGAGAGACACATATAAAGA 639  
 600 CAACCTCATCAGCCTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAGCC 659  
 640 CAACCTCATCAGCCTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAGCC 699  
 660 TGAGTACCAACACAGAGCGGTGATGCTCTCATGCAAGATTCTACGTGGTTGAGAGTATCT 719  
 700 TGAGTACC - ACCAGAGCGGTGATGCTCTCATGCAAGA - TTCTAGTGGTTGAGAGTATCT 757  
 720 TCTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAAAGACTTTCGTTTCA 779  
 758 TCTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAAAGACTTTCGTTTCA 817  
 780 AGACCTATG - CACCTGTGCTTCGGCTACTTCC - GGGAGCATTTTGTATCCGG - --CC 834  
 818 AGACCTATGCTGCTTGGCTTCCGCTACTTCCGGGAGGATTTTGGGAATCCGGCCCC 877  
 835 CGATGATTAATCTTATCCCTCTGCA - GTGAGCGCTGATTTGAATCTGTAGCTC - TGGA 892  
 878 CATGATTAATCTTATCCCTCTGCAAGGAGCGCTGATTAATCTGTAGCTCTGGA 937  
 893 GCTAGTGGTCCCTTATCTA - TGTGTCAGCAGCATGATGATTCATTAATGAAGAGTCC 950  
 938 GCTAGGCTCCCTTATCAAGGGGGCCAGCCGACCATGGGTCTTTTATTAAGACCGGCC 997  
 951 AACATAAGAGCGGAA - ---TTTCTGCAAGCTGCTTCCAGGATACATG - -AACCT 1005  
 998 CACTTAAAGGGGGAATTTTGGGAAGCTTGTCTTCCGGAACTTAACCTGGAACCT 1057  
 1006 CAACAGAACCTCGGACTTGTCT 1029  
 1058 CACCCCAAACTCGGAATTTGTT 1081

RESULT 11  
 BM455340

LOCUS BM455340 1046 bp mRNA linear EST 05-FEB-2002  
 DEFINITION AGENCOURT\_6407203 NIH\_MGC\_85 Homo sapiens cdna clone IMAGE:5500311  
 5', mRNA sequence.  
 ACCESSION BM455340  
 VERSION BM455340.1 GI:18504380  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1046)  
 NIH-MGC http://mgi.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE

JOURNAL COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb-re@mail.nih.gov  
 Tissue Procurement: Lou Staudt  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation (LLNL)  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM12135 row: a column: 16  
 High quality sequence stop: 670.

FEATURES

source

Location/Qualifiers

1..1046  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5500311"  
 /clone\_lib="NIH\_MGC\_85"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lymph; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 222 a 277 c 301 g 246 t

ORIGIN

Query Match 21.0%; Score 779.2; DB 13; Length 1046;  
 Best Local Similarity 99.1%; Pred. No. 9.2e-169;  
 Matches 811; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

Qy 1 ATTAACAGCGCGTGTAGGAAGACGAGAGAGGGCGGTTCCTCTTTGGGACTTTTCA 60  
 Db 9 ATTAACAGCGCGTGTAGGAAGACGAGAGAGGGCGGTTCCTCTTTGGGACTTTTCA 68  
 Qy 61 TGCCTCGTTTTTTTTCAGATGTGGCTTGGTCTGGGCGCAAGGTCCAGCAGCCAGCTTA 120  
 Db 69 TGCCTCGTTTTTTTTCAGATGTGGCTTGGTCTGGGCGCAAGGTCCAGCAGCCAGCTTA 128  
 Qy 121 AGCTTACTCTCTGTGAAAGGGGAAAGTATCCCTGTGAAAGCGTTAAACTTTGGAG 180  
 Db 129 AGCTTACTCTCTGTGAAAGGGGAAAGTATCCCTGTGAAAGCGTTAAACTTTGGAG 188  
 Qy 181 GGGGTGCGGGACGTGAGTCTTCCCATCCAGGCGAATGTTGGCTTGGCTTGGCTTCC 240  
 Db 189 GGGGTGCGGGACGTGAGTCTTCCCATCCAGGCGAATGTTGGCTTGGCTTGGCTTCC 248  
 Qy 241 AGGAGCGGCTCGACGTCTGAGGGAGG - CCCGAGGGGGGGGGAGGTGCCCCACAGA 299  
 Db 249 AGGAGCGGCTCGACGTCTGAGGGAGGCCCCCGAGGGGGGGAGGTGCCCCACAGA 308  
 Qy 300 ACGGGGTTCTGTAAGAGAGCGTTGGGAAGATTCCATCCGAGAGAGAGAAACCGGAT 359  
 Db 309 ACGGGGTTCTGTAAGAGAGCGTTGGGAAGATTCCATCCGAGAGAGAGAAACCGGAT 368  
 Qy 360 TGAAGAGAGCGGCGCTGAGGGGAGGGGCTGCTAAGATGGCGTCCGCTCTCCG 419  
 Db 369 TGAAGAGAGCGGCGCTGAGGGGAGGGGCTGCTAAGATGGCGTCCGCTCTCCG 428  
 Qy 420 GGCGTCTCTTCGGTGGTTTTTTCATCCTTTGATCCCGCGGTCCCTTCTCTTACCTGT 479  
 Db 429 GGCGTCTCTTCGGTGGTTTTTTCATCCTTTGATCCCGCGGTCCCTTCTCTTACCTGT 488  
 Qy 480 CCTCAGCATCTGGAATCAAGAGAGCCATGGCATCTGAGGTGCTTATGCCCTCTGGCATGC 539  
 Db 489 CCTCAGCATCTGGAATCAAGAGAGCCATGGCATCTGAGGTGCTTATGCCCTCTGGCATGC 548  
 Qy 540 CCATCAAGAAAATAGGCATAGAAGTGTGATTCCTCAGGAGAGACACATATAAAGA 599  
 Db 549 CCATCAAGAAAATAGGCATAGAAGTGTGATTCCTCAGGAGAGACACATATAAAGA 608  
 Qy 600 CAACCTCATCAGCCTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAGCC 659

```

|||||
Db 609 CACCTCATCAGCCTTGAAGGTGCCATCAGTTAGGCATACCCACATGTGGGAGCC 668
QY 660 TGAGTACCAACAGAGCGTGATGTCTCATGCAAGATTTCACGTGGTTGAGAGTATCT 719
Db 669 TGAGTACCAACAGAGCGTGATGTCTCATGCAAGATTTCACGTGGTTGAGAGTATCT 728
QY 720 TC-TTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAAATGACITTCGTTTC 778
Db 729 TCTTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAAATGACITTCGTTTC 788
QY 779 AGACCTATGCACCTCTTG-CCTTCGGCTACTTCCGGG 815
Db 789 AGACCTATGCACCTCTTGCCCTTCGCTACTTCCGGG 826

RESULT 12
BI090667
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11027 row: i column: 06
High quality sequence stop: 793.
FEATURES
source
Location/Qualifiers
1..864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4998605"
/clone.lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 214 a 196 c 184 g 269 t 1 others
ORIGIN
Query Match 20.8%; Score 772.8; DB 13; Length 864;
Best Local Similarity 97.2%; Pred. No. 2.7e-167;
Matches 829; Conservative 0; Mismatches 18; Indels 6; Gaps 4;

QY 2604 CACAGCAGACAAATTTCAAGCCATTTTCAGATCAGAACTCCAGAGTGTTCAGCAAGATG 2663
Db 1 CACAGCAGACAAATTTCAAGCCATTTTCAGATCAGAACTCCAGAGTGTTCAGCAAGATG 60
QY 2664 CATTATCGTAGATTCCTCCAGAAAGCCATGTTTATGAGAGAGAGTAGTGATG 2723
Db 61 CATTATCGTAGATTCCTCCAGAAAGCCATGTTTATGAGAGAGAGTAGTGATG 120
QY 2724 CTCTGCCAGAACAGCTCTCTTTAAACTCCCTCTCTTGATGAATTTCTTAAGGCTGA 2783
Db 121 CTCTGCCAGAACAGCTCTCTTTAAACTCCCTCTCTTGATGAATTTCTTAAGGCTGA 180

```

```

QY 2784 AGSAATGAAGAGAGTGGGACATGGGTAATCTTTATCCCTTTTGTAAAAACAGAGGCAG 2843
Db 181 AGSAATGAAGAGAGTGGGACATGGGTAATCTTTATCCCTTTTGTAAAAACAGAGGCAG 240
QY 2844 CCATGGGCTGGGAGATCATAGCCCTTCCTAGGAGAAATCCTGTTCACGTCCAGGCTATAG 2903
Db 241 CCATGGGCTGGGAGATCATAGCCCTTCCTAGGAGAAATCCTGTTCACGTCCAGGCTATAG 300
QY 2904 TAATATTACTATTTTGCATTTGAATATATCTCGTTGTGTTTCTTAATGGAAGACT 2963
Db 301 TAATATTACTATTTTGCATTTGAATATATCTCGTTGTGTTTCTTAATGGAAGACT 360
QY 2964 TACCAATGAATTTTAGATCATCTCCAGAGGAGATTTTTTGTCTCTCTCATCTTTTC 3023
Db 361 TACCAATGAATTTTAGATCATCTCCAGAGGAGATTTCTTTTGTCTCTCTCATCTTTTC 420
QY 3024 CACAGTGTTCCTCCTGTTTGTGGAGCTAAGTAAAGAGGGGACACTCTGTCTGTTTAAAC 3083
Db 421 CACAGTGTTCCTCCTGTTTGTGGAGCTAAGTAAAGAGGGGACACTCTGTCTGTTTAAAC 480
QY 3084 AGACAGTCCATATCTGTGAGGCCAGCAATATTTTCTTAAACTCATGGGAGACAGAGA 3143
Db 481 AGACAGTCCATATCTGTGAGGCCAGCAATATTTTCTTAAACTCATGGGAGACAGAGA 540
QY 3144 TTCTTGCCCTTGTGAGGTCATCTGTGCCATATGTCTTACCCCTCTGTCTTCATGCAGG 3203
Db 541 TTCTTGCCCTTGTGAGGTCATCTGTGCCATATGTCTTACCCCTCTGTCTTCATGCAGG 600
QY 3204 GAAGTTGGAATGGGGGCTACATATGCCCTCTCCTCCCGTCTACAAGAGT-TGTGGTTT 3262
Db 601 GAAGTTGGAATGGGGGCTACATATGCCCTCTCCTCCCGTCTACAAGAGTATGTGGTTT 660
QY 3263 TCATCTGATCTTCCACTCTTGTGAGGGAAGAGGGGCTGTGTATCTCAGGCAGAT 3322
Db 661 TCATCTGATCTTCCACTCTTGTGAGGGAAGAGGGGCTGTGTATCTCAGGCAGAT 720
QY 3323 GTTGAATTCCTGTCTATCCCTTCTATCCACCCCTGTCTGATATATGTTAGCCCAT 3382
Db 721 GTTGAATTCCTGTCTATCCCTTCTATCCACCCCTGTCTGATATATGTTAGCCCAT 779
QY 3383 ACCCAATAA---CTGCTATATAGACACCCCGGAGGCTTTCGGCTGCTGCTTT 3439
Db 780 ACCCAATAA---CTGCTATATAGACACCCCGGAGGCTTTCGGCTGCTGCTTT 838
QY 3440 GCTGCCATGTTT 3452
Db 839 GCTGCCATGTTT 851

RESULT 13
AUI30897
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 824)
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saiko, K., Kawai, Y.,
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isoqai, T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975

```



```
|||||
Db 303 ACGGGGTTCTGTAAAGAGAGCTGGGAGATTGCAATTCGAGAGAGGAGGAAACCGGAT 362
QY 360 TGAAGAGAGCCAGGCGCTGAGGGGAGGGGCTGCTAAGATGGGCTGGCCCTCTCCG 419
Db 363 TGAAGAGAGCCAGGCGCTGAGGGGAGGGGCTGCTAAGATGGGCTGGCCCTCTCCG 422
QY 420 GCGCGCTCTCTCGGTCGGTTTTTCATCCTTTTCATCCCGGGTCCCTTCCTGTACCTTGT 479
Db 423 GCGCGCTCTCTCGGTCGGTTTTTCATCCTTTTCATCCTCCCGGGTCCCTTCCTGTACCTTGT 482
QY 480 CCT---CAGCATCTGGAATCAAGAGAGCCCATGGCATCTGAGGTGCCTTATGCTCTGGCA 536
Db 483 CCTCAGCAGCATCTGGAATCAAGAGAGCCCATGGCATCTGAGGTGCCTTATGCTCTGGCA 542
QY 537 TGCCCATCAAGAAAATAGGCCATAGAGTGTGATTTCCTCAGGAGACAAACATATAAAA 596
Db 543 TGCCCATCAAGAAAATAGGCCATAGAGTGTGATTTCCTCAGGAGACAAACATATAAAA 602
QY 597 AGACAACCTCATCAGCCTTGAAAGGTGCCATCCAGTTAGGCATTACCCACACTGTGGGA 656
Db 603 AGACAACCTCATCAGCCTTGAAAGGTGCCATCCAGTTAGGCATTACCCACACTGTGGGA 662
QY 657 GCCTGAGTACCAACACAGAGCGTGATGTCTCATGCAAGATTTCTACGTGTGAGAGTA 716
Db 663 GCCTGAGTACCAACACAGAGCGTGATGTCTCATGCG-AGATTTCTACGTGTGAGAGTA 721
QY 717 TCCTCTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAATGACTTCGTT 776
Db 722 TCCTCTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCACTACAATGACTTCGTT 781
QY 777 TC-AAGACCTATGCACC--TGTTGCCCTTCGGTACTTCC--GGGAGCTATTGATATCCGG 832
Db 782 TCAAGACCTATGCACCCTTGTTGCCCTTCGGTACTTCCGGGAGCTATTGATATCCGG 841
QY 833 CCGATGA--TTACTTGTATTCCCTCTGCAGTG-AGCCGCTGATTGAATCTGTACTCT 889
Db 842 GCCCAAGATTAATTGGAATTCCTCTGCAATGAACCCCTGGATGAATCTGTAGTTTT 901
QY 890 GG 891
Db 902 GG 903

RESULT 15
AL522834
LOCUS
DEFINITION AL522834 LTI_NFL004_NBC2 Homo sapiens cDNA clone EST 13-FEB-2001
prime, mRNA sequence.
ACCESSION AL522834
VERSION AL522834.1 GI:12786327
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..783
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB009YB17"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

FEATURES
source
```

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 199 a 167 c 163 g 253 t  
ORIGIN

Query Match 20.5%; Score 760.4; DB 9; Length 783;  
Best Local Similarity 98.9%; Pred. No. 1.9e-164;  
Matches 775; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

```
QY 2467 TTAATCCCTACCTTAGTCTTTCTCTCTGACTCTCTGGAAGAAATACTCTGTAATCTCTGTA 2526
Db 1 TTAATCCCTACCTTAGTCTTTCTCTCTGACTCTCTGGAAGAAATACTCTGTAATCTCTGTA 60
QY 2527 AAGTTTTTTGGGGGATAAGGGTGTGTTAAACACCTCCAGCTTCTCTCTCTCTCTCTTTT 2586
Db 61 AAGTTTTTTGGGGGATAAGGGTGTGTTAAACACCTCCAGCTTCTCTCTCTCTCTCTTTT 120
QY 2587 CTGAAAAAAGGAAAAAGCACACAGACACAAATTTCAAGCCATTTTCAGATCAAGAACTCCA 2646
Db 121 TCTGAAAAAAGGAAAAA-CACACAGACACAAATTTCAAGCCATTTTCAGATCAAGAACTCCA 179
QY 2647 GAAGTGTTCACAGATGCGCTATTCGTAGAGTTCCTCTCAGAAGAGCCATGGTGTATGAA 2706
Db 180 GAAGTGTTCACAGATGCGCTATTCGTAGAGTTCCTCTCAGAAGAGCCATGGTGTATGAA 239
QY 2707 GAGAAGAGTAGTGATGCTCTGCCAGAAGCAGCTCTCTTTAAACCTCCCTCTCTCTGAT 2766
Db 240 GAGAAGAGTAGTGATGCTCTGCCAGAAGCAGCTCTCTTTAAACCTCCCTCTCTGAT 299
QY 2767 GAATTTCTTAAGGCTGAAGGAATGAAGAGAGTGGGACATGGGGTAATCTTTATCCCTTTT 2826
Db 300 GAATTTCTTAAGGCTGAAGGAATGAAGAGAGTGGGACATGGGGTAATCTTTATCCCTTTT 359
QY 2827 GTTAAACAGAGGAGCCATGGCTGGGAGATCATAGCCCTCTCTAGGAGAGTCCCTGT 2886
Db 360 GTTAAACAGAGGAGCCATGGCTGGGAGATCATAGCCCTCTCTAGGAGAGTCCCTGT 419
QY 2887 TCACCTGCCAGGCTATAGTAATTAATTAATTTGCAATTTGAAATATATCTTGGTTGTTT 2946
Db 420 TCACCTGCCAGGCTATAGTAATTAATTAATTTGCAATTTGAAATATATCTTGGTTGTTT 479
QY 2947 TTCTAAATCTGAAGACTTACCAAAATGAATTTAGATCATCTCCAGAGGAGATTTTTTTT 3006
Db 480 TTCTAAATCTGAAGACTTACCAAAATGAATTTAGATCATCTCCAGAGGAGATTTTTTTT 539
QY 3007 GCTCTCTCATCTTTTCCAAACAGTGTCTCTGTTGTTGGAGCTAAGGTAAGAGGGGAC 3066
Db 540 GCTCTCTCATCTTTTCCAAACAGTGTCTCTGTTGTTGGAGCTAAGGTAAGAGGGGAC 599
QY 3067 ACTTCTGTCTGTTTAAACACAGACAGTCCATATCTGTGAGGCCAGCAAAATATTTCTTAACT 3126
Db 600 ACTTCTGTCTGTTTAAACACAGACAGTCCATATCTGTGAGGCCAGCAAAATATTTCTTAACT 659
QY 3127 CATGGGAGACAGCATTTCTTGCCTTGGTGGGTCAATGCTGTGTCATATGCTTACCC 3186
Db 660 CATGGGAGACAGCATTTCTTGCCTTGGTGGGTCAATGCTGTGTCATATGCTTACCC 719
QY 3187 CCCTGCTCTCATGCGGGGAGTTGAAATGGGGCTACATATGCGCTCTCCCTCCCGGTCT 3246
Db 720 CCCTGCTCTCATGCGGGGAGTTGAAATGGGGCTACATATGCGCTCTCCCTCCCGGTCT 779
QY 3247 ACA 3250
Db 780 A 783
```

Thu Jan 9 08:54:08 2003

us-10-003-354-3.rst

Page 13

Search completed: January 8, 2003, 18:06:40  
Job time : 4911 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 12:54:40 : Search time 9375 Seconds  
(without alignments)  
11526.263 Million cell updates/sec

Title: US-10-003-354-3  
Perfect score: 3713  
Sequence: 1 attaacagcgctggttagg.....aaacttaatgagttattta 3713

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_mu.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rnd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3713	100.0	3713	9	HSU78575	U78575 Human 68 kD
2	3374.8	90.9	3637	9	BC007833	BC007833 Homo sapi
3	3038.8	81.8	104531	9	HS0303A1	AL096800 Human DNA
4	3037.6	81.8	4262	9	BC028580	BC028580 Homo sapi
5	3037.2	81.8	182895	2	AC011078	AC011078 Homo sapi
6	2081.2	56.1	112804	9	AC006476	AC006476 Homo sapi
7	2047.2	55.1	225432	9	AF027390	AF027390 Homo sapi
8	2014	54.2	2133	9	HSU78576	U78576 Human 68 kD
9	1852	49.9	2010	9	HSU78577	U78577 Human 68 kD
10	1835	49.4	180189	9	AL365510	AL365510 Human DNA
11	1654.8	44.6	154526	2	AL592111	AL592111 Homo sapi
12	1654.8	44.6	159148	9	AL592424	AL592424 Human DNA
13	1618.2	43.6	184090	2	AC009564	AC009564 Homo sapi
14	1543.4	41.6	2553	10	BC031774	BC031774 Mus muscu
15	1537.2	41.4	2567	9	AK098097	AK098097 Homo sapi
16	1522.6	41.0	2403	10	D86177	D86177 Mouse mRNA
17	1259.8	33.9	1709	9	BC007005	BC007005 Homo sapi
18	1143	30.8	1143	9	HSU78578	U78578 Human 68 kD
19	772.8	20.8	2723	10	BC003763	BC003763 Mus muscu
20	735.4	19.8	3360	5	AY062923	AY062923 Xenopus l
21	683.6	18.4	4226	10	BC019138	BC019138 Mus muscu
22	680.4	18.3	2225	10	AB006916	AB006916 Mus muscu
23	674.4	18.2	5047	9	AB011161	AB011161 Homo sapi
24	633.4	17.1	63747	2	AC015754	AC015754 Homo sapi
25	610.8	16.5	630	9	HUWY158E12	D86176 Mouse mRNA
26	588.8	15.9	2378	10	BC034864	BC034864 Mus muscu
27	588.8	15.9	2414	10	BC048695	BC048695 Mus muscu
28	583.2	15.7	1651	10	AF048695	AF048695 Sequence 67
29	581.6	15.7	1623	6	A59517	A59517 Sequence 9
30	581.6	15.7	2764	9	HSSTM7	X92493 H.sapiens m
31	581.6	15.7	2764	9	HSSTM7	X92493 H.sapiens m
32	565.4	15.2	3318	9	BC030587	BC030587 Homo sapi
33	489.8	13.2	63747	2	AC015754	AC015754 Homo sapi
34	463.6	12.5	204857	2	AC084272	AC084272 Mus muscu
35	463.6	12.5	224447	10	AC087062	AC087062 Mus muscu
36	408	11.0	1899	6	AX333345	AX333345 Sequence
37	408	11.0	1899	9	HSU78581	U78581 Human type
38	357.4	9.6	3402	3	AY069204	AY069204 Drosophil
39	346.6	9.3	2980	3	AY058737	AY058737 Drosophil
40	346.6	9.3	85862	3	AC004433	AC004433 Drosophil
41	346.6	9.3	163012	3	AC008289	AC008289 Drosophil
42	346.6	9.3	171972	3	AC012388	AC012388 Drosophil
43	346.6	9.3	304247	3	AE003452	AE003452 Drosophil
44	343.4	9.2	2922	3	AF071417	AF071417 Drosophil
45	336.2	9.1	13653	2	AC018218	AC018218 Drosophil

ALIGNMENTS

RESULT 1  
HSU78575  
LOCUS  
DEFINITION  
Human 68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha  
mRNA, clone PIP5K1a1, complete cds.  
ACCESSION  
U78575  
VERSION  
U78575.1  
KEYWORDS  
GI:1743870  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
REFERENCE  
1 (bases 1 to 3713)  
AUTHORS  
Lolijens, J.C. and Anderson, R.A.  
TITLE  
Type I phosphatidylinositol-4-phosphate 5-kinases are distinct

HSU78575 3713 bp mRNA linear PRI 20-DEC-1996  
Human 68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha  
mRNA, clone PIP5K1a1, complete cds.

JOURNAL members of this novel lipid kinase family  
MEDLINE J. Biol. Chem. 271 (51), 32937-32943 (1996)  
PUBMED 97115834  
REFERENCE 8955136

## 2 (bases 1 to 3713)

AUTHORS Lofjens, J.C. and Anderson, R.A.

TITLE Direct Submission

JOURNAL Submitted (18-NOV-1996) Pharmacology, University of Wisconsin -  
Madison, 1300 University Ave., Madison, WI 53706, USA

## FEATURES

source

1. .3713

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="P1P5Ktal"

/cclone\_lib="human fetal brain 5'-stretch cDNA library

(Clontech # HL1149x)"

401..2050

/EC\_number="2.7.1.68"

/note="isoforms, possibly alternatively spliced, encoded

by GenBank Accession Numbers U78576 and U78577"

/codon\_start=1

/product="68 kDa type I phosphatidylinositol-4-phosphate

5-kinase alpha"

/protein\_id="AAC50910.1"

/db\_xref="GI:1743871"

/translation="MASASGSPSSVGFSDPAVPSCTLSSASGIKRPMASEVPIAS  
GMPKIKGHSVDSSGTYKKTSSALKAIGIQTHTVSLSTKPERDVLMDFFV  
VESIFPFGSLNLPDAHYNDFRVYAPVAFRPRELFGIRPDYLYSLCSEPLIEL  
CSSGASGSLFVSSDEFLIKTVQKHEAEFLQKLLPGVYMLNPNRLLPKFYLAC  
VOAGKNIIVVMNLLPRSKMHLKIDLGSTYKRASOKEREKPLTFPKDLDFLDQ  
IPDGLFADIMYNAICKLQRLDCLVQFKIMDSLMSIHNIHQREPLSETQYS  
VTRPAPOKALITAMESIOEARRGSTMETDDHMGIPARNKSGERLLYIGLID  
LQSYRFKKLHSHWKLVDGTVSVHRPGFAERFORFMCNTVFKPLPSPSKFE  
RSGSFRRAGSSNSCITYOPSVSHEHKAQVTTKAEVEPGVHLGRPDVLPQTPLEE  
ISEGSPIDPPSPVLPVGTQMLTSTTLEKVAESEFTH"

polya\_signal 3691..3696

polya\_site 3713

BASE COUNT 919 a 865 c 882 g 1047 t

## ORIGIN

Query Match 100.0%; Score 3713; DB 9; Length 3713;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3713; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTAACAGGCGGTGGTTAGAGAGCGGAGAGGGCGTTCCGTCCTTTGGGACTTTTCA 60  
DDB 1 ATTAACAGGCGGTGGTTAGAGAGCGGAGAGGGCGTTCCGTCCTTTGGGACTTTTCA 60  
QY 61 TGCCTCGTTTTTTTTCAGATCTGGCTTGGTCTGGCGCAAGGTCCACGAGCCACTTA 120  
DDB 61 TGCCTCGTTTTTTTTCAGATCTGGCTTGGTCTGGCGCAAGGTCCACGAGCCACTTA 120  
QY 121 AGCTTACTCTTCTGTGAAGGGGAAAGTATCCCTCTGTGAAGCGGTAAACTTGTGGAG 180  
DDB 121 AGCTTACTCTTCTGTGAAGGGGAAAGTATCCCTCTGTGAAGCGGTAAACTTGTGGAG 180  
QY 181 GGGGTGGGAGCGTAGTCTTCCCATCCAGGGCAATGGTGTGGCCCTTGAGCTGGTCC 240  
DDB 181 GGGGTGGGAGCGTAGTCTTCCCATCCAGGGCAATGGTGTGGCCCTTGAGCTGGTCC 240  
QY 241 AGGAGCGGCTGCAGCTGTCTGAGGAGCGCCGAGGGGGGGGGAGGTGGCCACAGAA 300  
DDB 241 AGGAGCGGCTGCAGCTGTCTGAGGAGCGCCGAGGGGGGGGGAGGTGGCCACAGAA 300  
QY 301 CGCGGGTCTGTAAAGAGACGTTGGGAAGATTCGATTCCGAGAAGGAAGAACCGGATT 360  
DDB 301 CGCGGGTCTGTAAAGAGACGTTGGGAAGATTCGATTCCGAGAAGGAAGAACCGGATT 360  
QY 361 GAAAGAGAGCCAGCGCTGAGGGGAGGGGGCTGCTAAGATGGCGCTCGCCCTCCCGG 420  
DDB 361 GAAAGAGAGCCAGCGCTGAGGGGAGGGGGCTGCTAAGATGGCGCTCGCCCTCCCGG 420  
QY 421 GCCGTGCTCTCGGTGGGTTTTTCATCCCTTGTATCCCGGGGTCCTTCTGTACCTTGTG 480

Db 1501 TGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAAGGGGAAAGGCTTCTGCTTTATAT 1560  
Qy 1561 TGGCATATTCACATCTTACAGTCTTACAGCTTTGTTAAGAGTTGGAGCACTCTTGAA 1620  
Db 1561 TGGCATATTCACATCTTACAGTCTTACAGCTTTGTTAAGAGTTGGAGCACTCTTGAA 1620  
Qy 1621 AGCCCTGTACATGACGGAGACACTGTCTCAGTGCATCGCCAGGCTTCTACGCTGAACG 1680  
Db 1621 AGCCCTGTACATGACGGAGACACTGTCTCAGTGCATCGCCAGGCTTCTACGCTGAACG 1680  
Qy 1681 GTTCCAGGCTTCATGTGCAACACAGATATTTAAGAGATTTCCCTTGAAAGCCTTCTCCTTC 1740  
Db 1681 GTTCCAGGCTTCATGTGCAACACAGATATTTAAGAGATTTCCCTTGAAAGCCTTCTCCTTC 1740  
Qy 1741 CAATAAGTTTCGGTCTGGCTCATCTTCTCTCGGGAGCAGGCTCCAGTGGCAACTCCTG 1800  
Db 1741 CAATAAGTTTCGGTCTGGCTCATCTTCTCTCGGGAGCAGGCTCCAGTGGCAACTCCTG 1800  
Qy 1801 CATTACTTACCAGCATCGGTCTCTGGGAAACACAAGGCACAAGTGACAACAAAGGCAGA 1860  
Db 1801 CATTACTTACCAGCATCGGTCTCTGGGAAACACAAGGCACAAGTGACAACAAAGGCAGA 1860  
Qy 1861 AGTGGAGCAGGCTTACCTTGGTGGTCTGATGTTTACCTCAGACTCCACCTTTGGA 1920  
Db 1861 AGTGGAGCAGGCTTACCTTGGTGGTCTGATGTTTACCTCAGACTCCACCTTTGGA 1920  
Qy 1921 GGAATCAGTGGGCTCGCCTATTCCTGACCCAGTTTCTACCTCTAGTTGGAGAC 1980  
Db 1921 GGAATCAGTGGGCTCGCCTATTCCTGACCCAGTTTCTACCTCTAGTTGGAGAC 1980  
Qy 1981 TTTGCAATGTAACTACAAGTACAACTTGGAAAGCTTGAAGTTGCAGAGTCAGAGTT 2040  
Db 1981 TTTGCAATGTAACTACAAGTACAACTTGGAAAGCTTGAAGTTGCAGAGTCAGAGTT 2040  
Qy 2041 CACCCATTAAGCGCAAGCCCTCAGAAGACCTTGAACAAAGTTCTGCCATCTCTGTGATCC 2100  
Db 2041 CACCCATTAAGCGCAAGCCCTCAGAAGACCTTGAACAAAGTTCTGCCATCTCTGTGATCC 2100  
Qy 2101 CAAGATGTACGCCCTTGGCCAGCAATGCTGAATTTTCTACTTGGTCAATCAAAAAG 2160  
Db 2101 CAAGATGTACGCCCTTGGCCAGCAATGCTGAATTTTCTACTTGGTCAATCAAAAAG 2160  
Qy 2161 GAGTGTAAAGTAGAGGGAGCTGCTCCCTCATCTTCTTCCCTGGAAGAACTCTCTC 2220  
Db 2161 GAGTGTAAAGTAGAGGGAGCTGCTCCCTCATCTTCTTCCCTGGAAGAACTCTCTC 2220  
Qy 2221 TCCCTCTCTTCCCTCATGAATGGCCCTTAGTGCCCTCAGAGAGTTGAGACCGCAGCATCC 2280  
Db 2221 TCCCTCTCTTCCCTCATGAATGGCCCTTAGTGCCCTCAGAGAGTTGAGACCGCAGCATCC 2280  
Qy 2281 CCTCCACTCAGAGTTGGGTGGTACGGAATTTCAACTTGGCCAGCCCTTGGCTCCACTAT 2340  
Db 2281 CCTCCACTCAGAGTTGGGTGGTACGGAATTTCAACTTGGCCAGCCCTTGGCTCCACTAT 2340  
Qy 2341 TGAATTTTTTTCAGACCCCAATCTTCATGCTGGAATGGATTTGCTGGACTTGGCAGCT 2400  
Db 2341 TGAATTTTTTTCAGACCCCAATCTTCATGCTGGAATGGATTTGCTGGACTTGGCAGCT 2400  
Qy 2401 TTTCTTCCCTCTGCTTTGACTAGGAACCGGACTCTTAATTTCTCAGGACAGACTAGCT 2460  
Db 2401 TTTCTTCCCTCTGCTTTGACTAGGAACCGGACTCTTAATTTCTCAGGACAGACTAGCT 2460  
Qy 2461 GGCACATATCCCTACCTAGTTCTTCTCTCTGACTCTCGGGAAGATACCTCCTGTAATC 2520  
Db 2461 GGCACATATCCCTACCTAGTTCTTCTCTCTGACTCTCGGGAAGATACCTCCTGTAATC 2520  
Qy 2521 TCTGTAAGGTTTTTGGGGATAGGTTTAAACCACCTCCAGCTTTCTTCTCTTTT 2580  
Db 2521 TCTGTAAGGTTTTTGGGGATAGGTTTAAACCACCTCCAGCTTTCTTCTCTTTT 2580  
Qy 2581 TTTTCTGAAAAAGGAAAAAGCACAGCACACAAATTTCAAGCCATTTTCAGATCAGA 2640  
Db 2581 TTTTCTGAAAAAGGAAAAAGCACAGCACACAAATTTCAAGCCATTTTCAGATCAGA 2640

Qy 2641 ACTCCAGAAGTGTTCACAAAGATGCCCTATTCGTAGAGTTCCCTCAGAAGAGCCATGGTGT 2700  
Db 2641 ACTCCAGAAGTGTTCACAAAGATGCCCTATTCGTAGAGTTCCCTCAGAAGAGCCATGGTGT 2700  
Qy 2701 TATGAAGAAGAAGTAGTGTGCTCTGCCAAGCAGCTCCTCTTAAACTCCTCCCTCT 2760  
Db 2701 TATGAAGAAGAAGTAGTGTGCTCTGCCAAGCAGCTCCTCTTAAACTCCTCCCTCT 2760  
Qy 2761 CTTGATGAATTTCTTAAGGCTGAAGAAATGAAGAGATGGGACATGGGTAATCTTTATC 2820  
Db 2761 CTTGATGAATTTCTTAAGGCTGAAGAAATGAAGAGATGGGACATGGGTAATCTTTATC 2820  
Qy 2821 CCTTTTGTAAACAGGAGGAGCCATGGCTGGGAGATCATAGCCCTTCTTAGGCAGAA 2880  
Db 2821 CCTTTTGTAAACAGGAGGAGCCATGGCTGGGAGATCATAGCCCTTCTTAGGCAGAA 2880  
Qy 2881 TCCCTGTTCACTGCCAGGCTATAGTAATTTACTATTTTGAATTTGAAATATATCTTGG 2940  
Db 2881 TCCCTGTTCACTGCCAGGCTATAGTAATTTACTATTTTGAATTTGAAATATATCTTGG 2940  
Qy 2941 TTGTTTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATTTCTCCAGAGAGATT 3000  
Db 2941 TTGTTTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATTTCTCCAGAGAGATT 3000  
Qy 3001 TTTTGTGCTCTCTCATCTTTTCCAAAGTGTTCCTCTGTTGTGGAGCTAAGGTAAAGA 3060  
Db 3001 TTTTGTGCTCTCTCATCTTTTCCAAAGTGTTCCTCTGTTGTGGAGCTAAGGTAAAGA 3060  
Qy 3061 GGGGACACTTCTGCTGTTTAAACAGACAGTCCATATCTGTGAGGCCAGCAATATTTCT 3120  
Db 3061 GGGGACACTTCTGCTGTTTAAACAGACAGTCCATATCTGTGAGGCCAGCAATATTTCT 3120  
Qy 3121 TAAACTCATGGGGAGACAGCAATTTCTGCCCTTGGTGGAGTTCATTCCTGCCATATGTC 3180  
Db 3121 TAAACTCATGGGGAGACAGCAATTTCTGCCCTTGGTGGAGTTCATTCCTGCCATATGTC 3180  
Qy 3181 CTACCCCTCTCTCTCATGCAAGGAGTTGGAATGGGGGCTACATATGCCCTCTCCTCC 3240  
Db 3181 CTACCCCTCTCTCTCATGCAAGGAGTTGGAATGGGGGCTACATATGCCCTCTCCTCC 3240  
Qy 3241 CCGTCTACAAGAGTTGTGTTTCCATCTGATCCTTCCACCTCTTGTGAGGGGAAAGGG 3300  
Db 3241 CCGTCTACAAGAGTTGTGTTTCCATCTGATCCTTCCACCTCTTGTGAGGGGAAAGGG 3300  
Qy 3301 GGCTGTATCTCAGGAGATTTGTAATTCCTGTTCTATCCCTCTCTATCCCACTCTG 3360  
Db 3301 GGCTGTATCTCAGGAGATTTGTAATTCCTGTTCTATCCCTCTCTATCCCACTCTG 3360  
Qy 3361 CTTGATTAATATGTAGCCCATACCCCAATTAACCTGCTATATAGACACCCCAAGCAG 3420  
Db 3361 CTTGATTAATATGTAGCCCATACCCCAATTAACCTGCTATATAGACACCCCAAGCAG 3420  
Qy 3421 TTTCTGGCTGCTGCTTTGCTGCCATCTTTTACAAAGAGGAAAGATTTCTGCTATT 3480  
Db 3421 TTTCTGGCTGCTGCTTTGCTGCCATCTTTTACAAAGAGGAAAGATTTCTGCTATT 3480  
Qy 3481 TTTTTCATAATTTACTATTTATGATGATTTTAAAGTGTATTTAAGGACAGATTTCTG 3540  
Db 3481 TTTTTCATAATTTACTATTTATGATGATTTTAAAGTGTATTTAAGGACAGATTTCTG 3540  
Qy 3541 TTAGGGTGGGAGGAATATTTGAGGAGGCTGGGCTTAGGGAAGGAATGGGGAAGC 3600  
Db 3541 TTAGGGTGGGAGGAATATTTGAGGAGGCTGGGCTTAGGGAAGGAATGGGGAAGC 3600  
Qy 3601 AACATTTTATTAAGTGTACTATTTGCTCTACTTTCTGTTCTGTTTTCAGAAATGGAATA 3660  
Db 3601 AACATTTTATTAAGTGTACTATTTGCTCTACTTTGCTGTTTTCAGAAATGGAATA 3660  
Qy 3661 CAATATAAAGTGATATATGTTTAAATGAATAACTTTAATGAGTTATTTA 3713  
Db 3661 CAATATAAAGTGATATATGTTTAAATGAATAACTTTAATGAGTTATTTA 3713

RESULT 2  
BC007833

LOCUS BC007833 3637 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, phosphatidylinositol-4-phosphate 5-kinase, type I,  
alpha, clone MGC:14107 IMAGE:4098514, mRNA, complete cds.

## ACCESSION

BC007833

## VERSION

BC007833.1 GI:14043746

## KEYWORDS

MGC.

## SOURCE

Homo sapiens.

## ORGANISM

Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 3637)  
Strausberg, R.  
Direct Submission  
Submitted (11-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

## COMMENT

Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Shewchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, J., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantropop, S., Thomas, P.J.,  
Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 20 Row: C Column: 6  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 1743872.

## FEATURES

## source

1..3637  
/organism="Homo sapiens"  
/db\_xref="LocusID:8394"  
/db\_xref="taxon:9606"  
/clone="MGC:14107 IMAGE:4098514"  
/tissue\_type="Muscle, rhabdomyosarcoma"  
/clone\_lib="NIH\_MGC\_17"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
454..1956  
/codon\_start=1  
/product="phosphatidylinositol-4-phosphate 5-kinase, type  
I, alpha"  
/protein\_id="AAH07833.1"  
/db\_xref="GI:14043747"

## CDS

1..3637  
/translation="MASASGSSSVGFSSFPDPAVPSTLSSASGIKRPMASEVPYAS  
GMPKIKHSRVSDEGTYKKTSSALKGAIOIGITHVGSLSKPKPRDVLMDFFVY  
VESIFPSEGSLPAHHYNDREPKTYAPAFRPRELFGIRPDYLSLSEPLIEL  
CSSGASGLFTVSSDDEIITVQKAEFLQKLLPGYKMLNPNRPTLLPFYGLYC  
VOAGKNIIVVMNLLPRSMHIKIDKLGSTYKRASQKEREKPLFFKDLDFLOD  
IPDGLFDADMIYNAKLCQDLQSFIMDYSLIMS IHNIDHAQREPLSSTQYS  
VDRPAPQKALYSTAMESIQEARRGCTMETDDHMGIPARNKSGRELLLYIGTIDI  
LQSYRFVKLHSHKALVHDGDTVSHRPGFVAFRFRMCNTVFKKIPCVHLRDPV  
LQPPPLEISEGSPIDPSPFSLVGLTQLMLTSTLEKLEVAESEFTH"  
906 a 834 c 868 g 1029 t

BASE COUNT  
ORIGIN

## Query Match

90.9%; Score 3374.8; DB 9; Length 3637;

Best Local Similarity 95.8%; Pred. No. 0;  
Matches 3558; Conservative 0; Mismatches 7; Indels 149; Gaps 3;

QY	1	ATTAAACAGGCGCTGGTTAGGAAGGAGCGGAGTTCCTCTCTTTGGGACATTTTCA	60
DB	53	ATTAAACAGGCGCTGGTTAGGAAGGAGCGGAGTTCCTCTCTTTGGGACATTTTCA	112
QY	61	TGCTTCGCTTTTCTTTCAGATGTGGTTCGGCGCAAGTCCAGACAGCAGCTTCA	120
DB	113	TGCTTCGCTTTTCTTTCAGATGTGGTTCGGCGCAAGTCCAGACAGCAGCTTCA	172
QY	121	ACGTTTACCTCTCTGTAAGGAGGAAAGTATCCCTCTGTAAGGCGTTCAGCTTGGAG	180
DB	173	ACGTTTACCTCTCTGTAAGGAGGAAAGTATCCCTCTGTAAGGCGTTCAGCTTGGAG	232
QY	181	GGGCGTGGGAGCGTGGTTCCTCCCATGCCAGGCGCAATGGTGTGGCCCTTGAGCTGGTCC	240
DB	233	GGGCGTGGGAGCGTGGTTCCTCCCATGCCAGGCGCAATGGTGTGGCCCTTGAGCTGGTCC	292
QY	241	AGGAGCGGCTCGAGTGTCTGAGGAGG - CCCGGAGGGGGGGGAGGTGGCCACAGA	299
DB	293	AGGAGCGGCTCGAGTGTCTGAGGAGGAGGCGGGGGGGGGAGGTGGCCACAGA	352
QY	300	ACGCGGCTCTGTAAGGAGCGTGGGAAGATTCCGATCCGAGAGAGAGCAAGCCGAT	359
DB	353	ACGCGGCTCTGTAAGGAGCGTGGGAAGATTCCGATCCGAGAGAGAGCAAGCCGAT	412
QY	360	TGAAGAGAGCCAGCGCTGAGGGGAGGGGCTGCTAAGATGGCGTGGCTCTCTCGG	419
DB	413	TGAAGAGAGCCAGCGCTGAGGGGAGGGGCTGCTAAGATGGCGTGGCTCTCTCGG	472
QY	420	GGCGCTCTGTTTCGGTTCGGTTCCTTTTCATCTCTTTCATCTCTCTCTCTCTCTCTCTCT	479
DB	473	GGCGCTCTGTTTCGGTTCGGTTCCTTTTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	532
QY	480	CCTCAGCATCTGGAATCAAGAGAGCCATGGGAGTGGGCTGAGGTGCTTATGCTCTGCGATCG	539
DB	533	CCTCAGCATCTGGAATCAAGAGAGCCATGGGAGTGGGCTGAGGTGCTTATGCTCTGCGATCG	592
QY	540	CCATCAAGAAATAGGCCATAGAGTGTGTTGATTCCTCAGGAGAGACAAATATAAAAGA	599
DB	593	CCATCAAGAAATAGGCCATAGAGTGTGTTGATTCCTCAGGAGAGACAAATATAAAAGA	652
QY	600	CAACCTCATCAGCTTGAAGAGTGCATCCAGTACGATGAGGATACCCACATGTGGGAGCC	659
DB	653	CAACCTCATCAGCTTGAAGAGTGCATCCAGTACGATGAGGATACCCACATGTGGGAGCC	712
QY	660	TGAGTACCAAAACAGAGCGTGTATGCTCTCATCAAGATTTCTACGTGTGAGAGTATCT	719
DB	713	TGAGTACCAAAACAGAGCGTGTATGCTCTCATCAAGATTTCTACGTGTGAGAGTATCT	772
QY	720	TCTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCAGTACATGATGCTTCTGTTTCA	779
DB	773	TCTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCATCAGTACATGATGCTTCTGTTTCA	832
QY	780	AGACCTATCAGCTGTGCTTCCGCTACTTCCGGAGGATTTGGTATCCGCGCCGATG	839
DB	833	AGACCTATCAGCTGTGCTTCCGCTACTTCCGGAGGATTTGGTATCCGCGCCGATG	892
QY	840	ATTACTTGTATTCCTCTCAGTGCAGCCCTGATTTGAATCTGTAGCTCTGGAGCTAGTG	899
DB	893	ATTACTTGTATTCCTCTCAGTGCAGCCCTGATTTGAATCTGTAGCTCTGGAGCTAGTG	952
QY	900	GTTCCCTATTTATGTGTCCAGGAGGATGAGTTCATTTAAGACAGTCCCAACATAAAG	959
DB	953	GTTCCCTATTTATGTGTCCAGGAGGATGAGTTCATTTAAGACAGTCCCAACATAAAG	1012
QY	960	AGCGGAAATTTCTGAGAGGCTCTCCAGGATACATGATGAACCTCAACAGAGACCCCTC	1019
DB	1013	AGCGGAAATTTCTGAGAGGCTCTCCAGGATACATGATGAACCTCAACAGAGACCCCTC	1072
QY	1020	GGACTTTGCTGCTAAATTTCTATGGACTACTGTGTCAGGAGGCTGGCAAGACATTC	1079



QY 3240 CCGCTCTACAAGAGTGTGGTTTTCCATCTGATCCTTCACACTCTTGTGTCAGGGGAAGAAGG 3299  
 Db 3145 CCGCTCTACAAGAGTGTGGTTTTCCATCTGATCCTTCACACTCTTGTGTCAGGGGAAGAAGG 3204  
 QY 3300 GGGCCCTGGTATCTCAGGCAGATGTTGAATTCCTGCTTATCCCTTCTATCCACCT 3359  
 Db 3205 GGGCCCTGGTATCTCAGGCAGATGTTGAATTCCTGCTTATCCCTTCTATCCACCT 3264  
 QY 3360 GCCTTGATAAATGTTAGCCCATACCCCAATAAAGTCTATATATAGACACCCCGGCA 3419  
 Db 3265 GCCTTGATAAATGTTAGCCCATACCCCAATAAAGTCTATATATAGACACCCCGGCA 3324  
 QY 3420 GTTCTGGCTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3479  
 Db 3325 GTTCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3384  
 QY 3480 TTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3539  
 Db 3385 TTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3444  
 QY 3540 GTTAGGGTGGGAGGAATATTTAGGGAGGCTGGTCTTAGGGAAGGAATGGGGAAG 3599  
 Db 3445 GTTAGGGTGGGAGGAATATTTAGGGAGGCTGGTCTTAGGGAAGGAATGGGGAAG 3504  
 QY 3600 CAACATTTTATAGTGTACTATTTGGCTCTACTTTGTATGTTGTCAGAAATGCAAT 3659  
 Db 3505 CAACATTTTATAGTGTACTATTTGGCTCTACTTTGTATGTTGTCAGAAATGCAAT 3564  
 QY 3660 ACAATATAAAGTATATGCTTTTAACTGTAATAAATTAATGAGTTATTTA 3713  
 Db 3565 ACAATATAAAGTATATGCTTTTAACTGTAATAAATTAATGAGTTATTTA 3618

RESULT 3  
 HSDJ303A1  
 LOCUS  
 DEFINITION Human DNA sequence from clone RPI-303A1 on chromosome 6. Contains an CXADR (cox sackie virus and adenovirus receptor) pseudogene, a 1-phosphatidylinositol-4-phosphate kinase (PTPINS(4)p-5-kinase, EC 2.7.1.68) pseudogene, the last exon of a putative new gene, ESTs, STSs, GSS and a putative CpG island, complete sequence.  
 ACCESSION AL096800  
 VERSION AL096800.20 GI:7242638  
 KEYWORDS HTG: CpG island; CXADR.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 104531)  
 Collier,R.  
 Direct Submission  
 Submitted (08-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Mar 14, 2000 this sequence version replaced gi:7159759.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>  
 RPI-303A1 is from the library RPI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RPI-303A1 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RPI-303A1 is at 1 in this sequence. The true left end of clone RPI-303A1 is at 104332 in this sequence. The true right end of clone RPI-126E20 is at 48305 in this sequence.  
 Location/Qualifiers  
 1. 104531  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RPI-303A1"  
 /clone\_lib="RPI-1"  
 91. 279  
 /note="L2 repeat: matches 2254. .2430 of consensus"  
 complement(297. 601)  
 /note="match: STS: Em:U63119"  
 347. 490  
 /note="72 copies 2 mer at 79% conserved"  
 349. 492  
 /note="36 copies 4 mer at 79% conserved"  
 496. 744  
 /note="TRIGGER2 repeat: matches 2461. .2715 of consensus"  
 802. 1417  
 /note="TRIGGER2 repeat: matches 126. .754 of consensus"  
 3163. 3450  
 /note="L1MD repeat: matches 997. .1286 of consensus"  
 3545. 3722  
 /note="L1MD repeat: matches 1363. .1531 of consensus"  
 3846. 9957  
 /note="L1PA6 repeat: matches 12. .6139 of consensus"  
 10628. 11157  
 /note="MER70B repeat: matches 123. .653 of consensus"  
 11307. 11467  
 /note="MER5A repeat: matches 4. .189 of consensus"  
 11526. 11920  
 /note="HAL1 repeat: matches 422. .831 of consensus"  
 12958. 13316  
 /note="MLT1A repeat: matches 2. .365 of consensus"  
 13627. 13929  
 /note="Alusx repeat: matches 1. .301 of consensus"  
 14627. 14935  
 /note="Alus repeat: matches 1. .313 of consensus"  
 15841. 16135  
 /note="Alusx repeat: matches 1. .302 of consensus"  
 16245. 16550  
 /note="MLT1B repeat: matches 1. .386 of consensus"  
 complement(16571. 17463)  
 /gene="dJ303A1.1"  
 complement(16571. 17463)  
 /gene="dJ303A1.1"  
 /note="dJ303A1.1 (CXADR (cox sackie virus and adenovirus receptor) pseudogene)  
 match: CDNAS: Em:AF109645 Em:U90715 Em:AF109646 Em:Y07593  
 Em:U90716 Em:AF124598 Em:Y11929 Em:Y10320  
 match: proteins: Tr:p97792 Tr:O09052 Tr:p78310"  
 /pseudo  
 /codon\_start=1  
 /evidence=not\_experimental  
 16701. 16766  
 /note="Alu repeat: matches 237. .302 of consensus"  
 18602. 18825  
 /note="56 copies 4 mer aaga 75% conserved"  
 18606. 18821  
 /note="8 copies 27 mer 56% conserved"

```
repeat_region 19495..19531
/notes="L2 repeat: matches 2617. .2653 of consensus"
repeat_region 19573..20005
/notes="MTC repeat: matches 1. .402 of consensus"
repeat_region 20263..20419
/notes="MER46C repeat: matches 2. .162 of consensus"
repeat_region 21023..21333
/notes="THE1B repeat: matches 1. .307 of consensus"
repeat_region 21335..27440
/notes="LIPAS repeat: matches 11. .6145 of consensus"
repeat_region 27446..27515
/notes="THE1B repeat: matches 294. .364 of consensus"
misc_feature 27921..28345
/notes="match: GSS: Em:AQ057927"
misc_feature 27956..28359
/notes="match: GSS: Em:AQ057921"
repeat_region 28549..28855
/notes="AluX repeat: matches 1. .309 of consensus"
repeat_region 30105..30229
/notes="FLAM.C repeat: matches 1. .124 of consensus"
misc_feature complement(30634..31065)
/notes="match: GSS: Em:AQ496311"
misc_feature complement(30644..31063)
/notes="match: GSS: Em:AQ593448"
repeat_region 30858..31142
/notes="AluJ repeat: matches 1. .294 of consensus"
misc_feature 31104..31403
/notes="match: GSS: Em:B63554"
repeat_region 32509..32782
/notes="AluJ repeat: matches 1. .284 of consensus"
repeat_region 33014..33181
/notes="FRAM repeat: matches 4. .170 of consensus"
repeat_region 33973..34274
/notes="AluX repeat: matches 1. .300 of consensus"
repeat_region 34836..35132
/notes="AluX repeat: matches 2. .296 of consensus"
repeat_region 35146..35195
/notes="25 copies 2 mer gt 84% conserved"
repeat_region 35160..35195
/notes="9 copies 4 mer gtgt 100% conserved"
repeat_region 35207..35258
/notes="26 copies 2 mer at 71% conserved"
repeat_region 35807..36268
/notes="L1A9 repeat: matches 5793. .6268 of consensus"
repeat_region 36269..36578
/notes="AluSg repeat: matches 1. .310 of consensus"
repeat_region 36579..36615
/notes="L1A9 repeat: matches 6268. .6303 of consensus"
repeat_region 36963..37278
/notes="AluX repeat: matches 1. .312 of consensus"
repeat_region 37957..38250
/notes="AluSg repeat: matches 1. .304 of consensus"
misc_feature complement(38069..38560)
/notes="match: GSS: Em:AQ592281"
repeat_region 38844..39046
/notes="L2 repeat: matches 2542. .2746 of consensus"
repeat_region 39048..39292
/notes="MER73 repeat: matches 376. .633 of consensus"
repeat_region 39293..39604
/notes="AluX repeat: matches 1. .300 of consensus"
repeat_region 39605..39959
/notes="MER73 repeat: matches 65. .376 of consensus"
repeat_region 39960..39995
/notes="18 copies 2 mer gt 100% conserved"
repeat_region 39998..40274
/notes="AluY repeat: matches 1. .277 of consensus"
repeat_region 40276..40348
/notes="MER73 repeat: matches 1. .74 of consensus"
repeat_region 40640..40778
/notes="LIME2 repeat: matches 5045. .5178 of consensus"
repeat_region 40788..40924
/notes="L1A10 repeat: matches 6180. .6316 of consensus"
repeat_region 40964..41130
```

```
repeat_region 41167..41424
/notes="AluJb repeat: matches 128. .296 of consensus"
repeat_region 41441..41608
/notes="LIME2 repeat: matches 5189. .5449 of consensus"
repeat_region 41609..41892
/notes="AluX repeat: matches 1. .293 of consensus"
repeat_region 41893..42166
/notes="L1A9 repeat: matches 6057. .6308 of consensus"
repeat_region 42909..43196
/notes="AluY repeat: matches 1. .287 of consensus"
repeat_region 43299..43875
/notes="LIME2 repeat: matches 5478. .6155 of consensus"
repeat_region 44188..44521
/notes="MUT1A repeat: matches 1. .363 of consensus"
repeat_region 44580..44725

Query Match 81.8%; Score 3038.8; DB 9; Length 104531;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 3420; Conservative 0; Mismatches 277; Indels 55; Gaps 16;

QY 2 TTACAGGCGGTGGTTAGGAAGGACGAGAGGGCGGTTCGCTCTCTTTGGGACTTTTCAT 61
Db 64268 TAAACAGGCGGTGGTTAGGAGGACAGAGAGGCGGTTCGCTCTCTCTGGGACTTTTCAC 64327
QY 62 GCCTCG-----TTTTTTTTTCAGATGTGGCTTGGGCGCAAGGTCCCAAGCAGCC 114
Db 64328 GCCTCGGTGTGTGTGAAAGGGGAAAGTATCCCTGTGTGAAAGCGTTAAACTT 174
QY 115 AGCTTAAGCTTACTTCTGTGAAAGGGGAAAGTATCCCTGTGTGAAAGCGTTAAACTT 174
Db 64388 AGCTTGAGCTTACTTCTGTGAAAGGGGAAAGTCTCTCTGAGGAAGCGGTTGAACCT 64447
QY 175 GTGAGAGGGGTGCGGACGTGAGTTCTTCCCATCCAGGCGAATGTGTGCCCTTGAGC 234
Db 64448 GTGAGGGCGTGCAGAGCGGCGTCTCCCATGCCAGGGAATGTGCGGCTTGAGC 64507
QY 235 TGTTCAGGAGCGGCTC---GAGCTGTCTGAGGAGG--CCCGAGGGGGGGGAGGTG 290
Db 64508 TGTTCAGAGCGCGCTCGATGACAGGTCTGAGGAGGCGCCAGAGCGCGGGAAGTG 64567
QY 291 GCCCAGAACGCGGTTCTGTAAGAGAGCGTTGGGAAGATTTCATCCGAGAAGAGAA 350
Db 64568 GCCCGAGACGCGGTTCTGTAAGAGACGTGGAGAAGATTTCATCCGAGAAGAGAA 64627
QY 351 GAACCGGATTGAAGAGAGAGCCCGCTGAGGAGGAGGGGCTCTCTAAGATGGCGTCG 410
Db 64628 GAGCGGATGGAAGAGACAGCCGCGCTGCCGGAAGGGGCTGCCAAGATGGCGTCG 64687
QY 411 CTCTCTCCGGGCGTCTGCTTCGGTGGTGTTCATCCTTTTGATCCCGCGGTCCCTTCCT 470
Db 64688 CTCTCTCCCAACCGTCTGCTGGGCTGTTTTCATCCTTTGATCCCGGGGCGCTTCCT 64747
QY 471 GTACCTTGCTCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGTGCTTTATGCT 530
Db 64748 GTACCGGCTCTCAGCATCTGGAATCTTGAGCCCAAGGCAATCTGAGTGCTTTATGCT 64807
QY 531 CTGCGATCCCATCAAGAAATAGGCCATAGAACTGTGTGATTCCTCAGGAGAGACAAT 590
Db 64808 CTGCGATCCCATCAAGAAATAGGCCATCGAGGTGTCGATTCCTCAGGAGAGACAAT 64867
QY 591 ATAAAAGACAACCTCATACGCTTGAAGGTGCCATCCAGTTAGGATATACCACTAGT 650
Db 64868 ATAAAAGACAACCTCAACAGCCTTGAAGGTGCCATCCAGTTAGGATATACCTAGT 64927
QY 651 TGGGAGCCTGAGTACCACCAAGAGCGTGTATGTCCTCATGCAAGATTTCTACCTGTTG 710
Db 64928 TGGGAGCCTGAGTACCACCAAGAGCGTGTATGTCCTCATGCAAGATTTCTACCTGTTG 64987
QY 711 AGAGTATCTCTTTTCCAGTGAAGGAGCAACCTTGACCCCTGCCTCATCAATGACT 770
Db 64988 AGAGTATCTCTTTCCAGTGAAGGAGCAACCTTGACCCCTGCCTCATCAATGACT 65047
```



QY 771 TTGCTTCAAGACCTATGACCTGTTGGCTTCGGCTACCTTCGGGAGCTATTGTTGATCC 830  
Db 65048 TCGTTTCAAGACCTATGACCTGTTGGCTTCGGCTACCTTCGGGAGCTATTGTTGATCC 65107  
QY 831 GGCCCGATGATTACTTGTATTCCTCTGAGTGAGCGCTGATTGAACCTGTAGCTG 890  
Db 65108 CGCCCGATGATTACTTGTGCTCCCTCTGAGTGAGCGCTGATTGAACCTGTAGCTG 65167  
QY 891 GAGCTAGTGGTTCCTTATCTATGTCTCAGGAGACGATGAGTTCAATTAAGACAGTCC 950  
Db 65168 GAGCTAGTGGTTCCTTATGTGTCTCAGGAGACGATGAGTTCAATTAAGACAGTCC 65227  
QY 951 AACAATAAGAGCGGGAATTTTCAGAGAGCTGCTCCAGGATACATGATGAACTCAAC 1010  
Db 65228 AACAATAAGAGCGGGAATTTTCAGAGAGCTGCTCCAGGATACATGATGAACTCAAC 65287  
QY 1011 AGAACCTCGGACTTGTGCTCAATCTATGAGTGTGCTGAGGAGCTGCTGAGGAGTGGCA 1070  
Db 65288 AGAACCTCGGACTTGTGCTCAATCTATGAGTGTGCTGAGGAGCTGCTGAGGAGTGGCA 65347  
QY 1071 AGAACCTCGGACTTGTGCTCAATCTATGAGTGTGCTGAGGAGCTGCTGAGGAGTGGCA 1130  
Db 65348 AGAACCTCGGACTTGTGCTCAATCTATGAGTGTGCTGAGGAGCTGCTGAGGAGTGGCA 65407  
QY 1131 AATATGACCTCAAGGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 1190  
Db 65408 AATATGACCTCAAGGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 65467  
QY 1191 CTCCTCCCACTTAAAGACCTAGACTTCTTACAGACATCCCTGATGCTTTTTTGG 1250  
Db 65468 CTCCTCCCACTTAAAGATCTAGACTTCTTACAGACATCCCTGATGCTTTTTTGG 65527  
QY 1251 ATGCTGACATGACAGCTCTCTGTAGAGCCCTGCAGCGTGAATTTGTTGCTGCAGA 1310  
Db 65528 ATGCTGACATGACAGCTCTCTGTAGAGCCCTGCAGCGTGAATTTGTTGCTGCAGA 65587  
QY 1311 GCTTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1370  
Db 65588 GCTTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 65647  
QY 1371 GAGAGCCCTTAAGACCTGAAACACAGTA---CTCAGTTGATCTGAAAGACCGGCCCC 1427  
Db 65648 GAGAGCCCTTAAGACCTGAAACACAGTA---CTCAGTTGATCTGAAAGACCGGCCCC 65707  
QY 1428 AAAAGGCTCTGTATTCAGAGCCATGGAATCCATCCAGGAGAGCTCGAGGGTGGTA 1487  
Db 65708 AAAAGGCTCTGTATTCAGAGCCATGGAATCCATCCAGGAGAGCTCGAGGGTGGTA 65767  
QY 1488 CCATGGAGACTGATGACCATATGGGTGGCATCCCTGCGGGAATAGTAAAGGGAAGGC 1547  
Db 65768 CCATGGAGCCGATGACCATATGGGTGGCATCCCTGCGGGAATAGTAAAGGGAAGGC 65827  
QY 1548 TTCTGCTTTATATGGCATATGATGATGATGATGATGATGATGATGATGATGATG 1607  
Db 65828 TTCTGCTTTATATGGCATATGATGATGATGATGATGATGATGATGATGATGATG 65887  
QY 1608 AGCACTCTGGAAGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1667  
Db 65888 AGCACTCTGGAAGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 65947  
QY 1668 TCTAGGCTGAAGCGTTCCAGCGCTTCAATGATGATGATGATGATGATGATGATGATG 1727  
Db 65948 TCTAGGCTGAAGCGTTCCAGCGCTTCAATGATGATGATGATGATGATGATGATGATG 66007  
QY 1728 AGCCTTCTCTTCAAAAAGTTTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1787  
Db 66008 AGCCTTCTCTTCAAAAAGTTTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 66067  
QY 1788 GTGGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1847  
Db 66068 GTGGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 66126  
QY 1848 CAACAAGGCGAGAGCTGGAGCGGCTTCACTTGGTGGTGGTGGTGGTGGTGGTGGT 1907

Db 66127 --ATAAGGTGCAAGTGGAGCGAGGTTCACCTTGGTGGTGGTGGTGGTGGTGGTGGT 66184  
QY 1908 C-----TCCACCTTTGGAGGAATCAGTGAGGCTGCCCTATTCTCGACCCCGAGTTTC 1960  
Db 66185 CTTCAAGATCCACCTTTGGAGGAATCAGTGAGGCTGCCCTATTCTACTGACCCAGCTTTC 66243  
QY 1961 TCACCTCTAGTTGGAGAGACTTTGCAATGCTAACGTACAAAGTACAACTTGGAAAAGCTT 2020  
Db 66244 TCACCTCTAGTTGGAGAGACTTTGCAATGCTAACGTACAAAGTACAACTTGGAAAAGCTT 66303  
QY 2021 GAAGTTGACAGTCAAGTTCACCTTTAAGCGCAAGCCTCAGAGAGCTCGAACAAGA 2080  
Db 66304 GAAGTTGACAGTCAAGTTCACCTTTAAGCGCAAGCCTCAGAGAGCTCGAACAAGA 66363  
QY 2081 TTCTGCCATCTCTGATGCCCAAGATGTCAGCCCTTGGCCCAAGCAATGCTGAATTTCTT 2140  
Db 66364 TTCTCTCTGATGCCCAAGATGTCAGCCCTTGGCCCAAGCAATGCTGAATTTCTT 66422  
QY 2141 CTACTTGGTCAATCAAAAAGAGTCTAATAGAAAGTGAAGGAGGAGCTCTCTCCATCTCT 2200  
Db 66423 CTACTTGGTCAATCAAAAAGAGTCTAATAGAAAGTGAAGGAGGAGGCTCTCTCTCT 66478  
QY 2201 TCCTGAAGAAGAACCTTCT 2260  
Db 66479 TCCTGAAGAAGAGCCT 66538  
QY 2261 AGTTTCAGGA-----CGCAGCATCCCTCCACAGAGTGGTGGTGGTGGTGGTGG 2306  
Db 66539 AGTTTCAGGACCACTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66598  
QY 2307 GATTTTCAACTGGCCCAACCTTTTGGCTCCACTTATTAATTTTTCAGACCCCACTCTT 2366  
Db 66599 GATTTTCAACTGGCTTAA---CTTTGGTTTCCACAACTGAATTTTTCAGACCCCACTCTT 66657  
QY 2367 CATGCTGAAATGGGATTTGCTGGACTTGGAGCTTTTCTCTCTCTCTCTCTCTCTCTCT 2426  
Db 66658 CATGCTGAAATGGGATTTGCTGGACTTGGAGCTTTTCTCTCTCTCTCTCTCTCTCTCT 66717  
QY 2427 ACCGACCTCTTAAATTTCTCTAGGACAGACTAGCTGGCAGATATCCCTACCTTACTTCT 2486  
Db 66718 ACTGACCTCTTAAATTTCTCTAGGACAGACTAGCTGGCAGATATCCCTACCTTACTTCT 66777  
QY 2487 TCTCTCTGACTCTCTGAAAGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2546  
Db 66778 TCTCTCTGACTCTCTGAAAGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66837  
QY 2547 GTGTTTAAACCACTCCGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2606  
Db 66838 GTGTTTAAACCACTCCGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66895  
QY 2607 ACAGCACAAATTTCAAGCCATTTTCAGATCAGAACTCCAGAGTGTGACAAAGATGCT 2666  
Db 66896 ACAGCACAAATTTCAAGCCATTTTCAGATCAGAACTCCAGAGTGTGACAAAGATGCT 66955  
QY 2667 ATTCGTAGAGTTCCTCAGAGAGCCATGGTGTGATGAGAGAGAGAGTGTGATGCTC 2726  
Db 66956 ATTCGTAGAGTTCCTCAGAGAGCCATGGTGTGATGAGAGAGAGTGTGATGCTC 67015  
QY 2727 TGCCAGAGAGAGCTCTCTTAACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2786  
Db 67016 TGCCAGAGAGAGCTCTCTTAACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 67075  
QY 2787 AATGAAGAGAGTGGGACATGGGGAATCTTTATCTCTCTCTCTCTCTCTCTCTCTCTCT 2846  
Db 67076 AATGAAGAGAGTGGGACATGGGGAATCTTTATCTCTCTCTCTCTCTCTCTCTCTCTCT 67135  
QY 2847 TGGGCTGGGACATGATGAGCCCTTCTAGGAGAGATCTGTTGTTTCTCTCTCTCTCTCTCT 2906  
Db 67136 TGGGCTGGGACATGATGAGCCCTTCTAGGAGAGATCTGTTGTTTCTCTCTCTCTCTCTCT 67195  
QY 2907 TTATTACTTATTTTGAATTTGAATATATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2966













DB 158002 TTTTATGTAATAAACTTTTATCAGTATTATA 158033  
|||||

## RESULT 6

AC006476 112804 bp DNA linear PRI 30-SEP-2000  
LOCUS Homo sapiens PAC clone RP4-764012 from 7, complete sequence.  
DEFINITION AC006476  
ACCESSION AC006476  
VERSION AC006476.3 GI:4926910  
KEYWORDS HTG.

## SOURCE

Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 112804)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
PUBMED 9847074

## REFERENCE

2 (bases 1 to 112804)  
Cordes, M., Wohlman, P. and Phillips, A.  
The sequence of Homo sapiens PAC clone RP4-764012  
Unpublished  
3 (bases 1 to 112804)  
Waterston, R.H.  
Direct Submission  
Submitted (30-JAN-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 112804)  
Waterston, R.H.  
Direct Submission  
Submitted (29-MAY-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 112804)  
Waterston, R.  
Direct Submission  
Submitted (30-SEP-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On May 29, 1999 this sequence version replaced gi:4337287.

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

mailto:egreen@nhgri.nih.gov , or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

This clone was derived from human PAC library RP4-764012, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc.

(<http://www.genomesystems.com>) or Research Genetics, Inc.

(<http://www.resgen.com>); or from Pieter de Jong.

## VECTOR: pCYPAC2

## NEIGHBORING SEQUENCE INFORMATION:

A tandem repeat, extending from base position 67681 to 71791, has been sorted to best possible assembly. Restriction digest information with bamhi, hindiii, and ecorv suggests this region is missing approximately 1400 base pairs of sequence.

The clone sequenced to the left is RP5-1058P19. Actual start of

this clone is at base position 1 of RP4-764012; actual end is at

112804 of RP4-764012.

## FEATURES

## source

1..112804 Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP4-764012"

/clone\_lib="RPCI-4"

1..65

/rpt\_family="Alu"

282..668

/rpt\_family="MaLR"

1575..1615

/rpt\_family="AT-rich"

1616..1897

/rpt\_family="Alu"

3392..3775

/rpt\_family="Retroviral"

4319..4868

/note="match to EST AI568539 (NID:94531913) tn40e08.xl"

5673..5921

/rpt\_family="MER1\_type"

8043..10028

/note="CpG island (%GC=68.4, o/e=0.90, #CpGs=214)"

8332..8364

/rpt\_family="A-rich"

9202..9330

/rpt\_family="(CCCCG)n"

9950..9970

/rpt\_family="GC-rich"

10090..10212

/rpt\_family="L1"

10986..11122

/rpt\_family="(TAGA)n"

11243..11360

/rpt\_family="(CAGA)n"

12015..12240

/rpt\_family="MER4-group"

12243..12853

/rpt\_family="MER4-group"

12881..12991

/rpt\_family="MER94"

12998..13524

/rpt\_family="L1"

16196..16425

/rpt\_family="Retroviral"

17256..17613

/rpt\_family="Retroviral"

18007..18149

/rpt\_family="MER2\_type"

18224..18522

/rpt\_family="MER2\_type"

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> , send

```
repeat_region 18715..18832
/rpt_family="L1"
repeat_region 19286..19373
/rpt_family="L2"
repeat_region 20391..20481
/rpt_family="MIR"
repeat_region 21808..22111
/rpt_family="Alu"
repeat_region 22147..22188
/rpt_family="At_rich"
repeat_region 22392..22555
/rpt_family="MaLR"
repeat_region 23127..23212
/rpt_family="MaLR"
repeat_region 24894..25083
/rpt_family="MER4-group"
repeat_region 25147..25237
/rpt_family="MER4-group"
repeat_region 25269..25598
/rpt_family="MER4-group"
repeat_region 25616..26607
/rpt_family="MER4-group"
repeat_region 26945..27063
/rpt_family="MER4-group"
repeat_region 27210..27502
/rpt_family="Alu"
repeat_region 27503..27768
/rpt_family="MER4-group"
repeat_region 28822..29089
/rpt_family="MER4-group"
repeat_region 29111..29368
/rpt_family="MER4-group?"
repeat_region 29377..29464
/rpt_family="(CATAT)n"
repeat_region 29748..29786
/rpt_family="At_rich"
repeat_region 29893..30466
/rpt_family="L1"
repeat_region 30524..30638
/rpt_family="(CATA)n"
repeat_region 30886..30997
/rpt_family="(CA)n"
repeat_region 31015..31125
/rpt_family="(CAATA)n"
repeat_region 31099..31230
/rpt_family="(CACAC)n"
repeat_region 32285..33146
/rpt_family="L1"
repeat_region 33240..33373
/rpt_family="MER2_type"
repeat_region 33380..33559
/rpt_family="L1"
misc_feature 35690..37077
note="CpG island (%C=63.0, o/e=0.90, #CpGs=106)"
repeat_region 37890..40107
/rpt_family="MER2_type"
repeat_region 42804..43109
/rpt_family="Alu"
repeat_region 44304..44361
/rpt_family="MIR"

Query Match 56.18; Score 2081.2; DB 9; Length 112804;
Best Local Similarity 75.3; Pred. No. 0;
Matches 3248; Conservative 0; Mismatches 403; Indels 665; Gaps 26;

Qy 20 GAAGACGAGAGGGGCTTCGCTTTGGGACTTTTCATGCTCGCTTTTTTTTTCAG 79
Db 98250 GGAGACGAGCGGGGGCTTCATCGCTGGGCTTCATGCCACGCTTTTCTCCAG 98309

Qy 80 ATGTGGCTGTGCTGGCGCAAGGTCCACAGCAGCAGCTTAAGCTTACTCTCTGTGAAA 139
Db 98310 GTGTGACTTGTC--GGACGCGGATCCGGCAGCAGCTGGGGCTTAGTCTCTGTGAAA 98367
```

```
Qy 140 GGGAAAGTATCCCTGTGGAAAGGGGTAAACTTGTGGAGGGGTGCGGACGTGAGTT 199
Db 98368 GAGTAAAGTGCCTCCTCGGAGAGCGGGTAAACTCGCAAGGGCTGCGAGCGGAGTT 98427
Qy 200 CTTCCCATGCCAGCGGAATGGTGTGGCTTTGAGCTGTGTCCAGGAGCGGCTCCACGCTGT 259
Db 98428 C-TCCCCAGGCCAGCGAATGGGCGGCCCTGAGCTGTCTA-GAGCGGGCTCCACGGCT 98485
Qy 260 CTGAGGAGCGCCGAGGGGCGGGAGGTGGCCACAGAACCGGATGAAAGAGCGCCAGCC 319
Db 98486 CTGAGGGGCGCGGGGCGGGAGGTGGCCCA---GACGGGGTCTGTGAAGGA 98542
Qy 320 GTTTGGGAAG--ATTGATTCCGAGAGAGAGAAACCGGATGAAAGAGCGCCAGCC 376
Db 98543 CTTGGGGAAGGAAGATTCCATTTCGAGAGAGAAAGCGGATCAAGAG-GCCAGGCC 98601
Qy 377 GCTGAGGGGAGGGGCTGTAAAGATGGCGTGGCCCTCCTCCGGGCCCTCGTCTTCGGTC 436
Db 98602 GCTGAGGGG--AGGGGCTGCCAAGATGGCGTGTCTCCTCAGGGCCTCCTCGGCGTC 98659
Qy 437 GGTTCATCCTTTGATCCCGCGGTCCCTTCCTGTACTTGTGCTCAGCATCTGGAATC 496
Db 98660 AGTTTCATCCTTTGATCCAGGGTGTCTTCCTGTACTGTCTCAGCATCTGGAATC 98719
Qy 497 AAGAGCCCATGGCATCTGAGTGCCTTATGCCTGTGCATGCCCATCAAGAAATAGCC 556
Db 98720 CAGAGCCCATGGCGTCTGAGTGCCTTGTGCCCTGGGCATACCCATCAAGAAATAGCC 98779
Qy 557 CATAGAAGTGTGATTTCCTCAGGAGAGACAACATATAAAAGACAACCTCATCAGCCTTG 616
Db 98780 CATCGAGTGTGATTTCCTCAGGAGAGACAACCTATAGAAAGACGACCTCATCAGCCTTG 98839
Qy 617 AAAGTGCCATCCAGTTAGGCATTACCCACACTGTGGGGAGCCTGAGTACCACACAGAG 676
Db 98840 AAATGTGCATCCAGTTGTGCATTACTCACACTGTGGGGAGCCTGCATCCACACAGAG 98899
Qy 677 CGTATGTCCCTCATGCAAGATTTCTACGTGTGGAGTACTTCTTTCCAGTGAAGGG 736
Db 98900 CGTCATGTCCCTCATGTAAGATTTCTCGTGTGGGAGATATCTTTCCAGTGGAGGG 98959
Qy 737 AGCAACCTGACCCCTGCTCATCATCAATGACTTTCTTTTCAAGACCTATGCACCTGT 796
Db 98960 AGCAACCTGACCCCTGCTCATCATCGAGTGTCTTTTCAAGACCTATGCACCTGT 99019
Qy 797 GCCTTCGCTACTTCGCGGAGCTATTGCTGTTCGGGCCGATGATTTACTTATTCCTC 856
Db 99020 GCCTTCGCTACTTCGAGGAGCTATTGCTGTCTGCCCTGATGATGACTTGTACTCCCTC 99079
Qy 857 TGCAGTGAGCGCTGATTGAACTCTGTAGCTCTGGAGCTAGTGTCCCTATTCATCTG 916
Db 99080 TGCAGTGAGCGCTGATTGAACTCAGTAGCTCTGGAGCTGATGGTTCCCTGTACATGTG 99139
Qy 917 TCCAGCGACGATGAGTTTCATTATTAAAGACAGTCCCAACATAAAGAGCGGAATTTCTGCAG 976
Db 99140 TCCATCCACATGATTCATTATTAAACAGTCCCAACATAAAGAGCGGAGTTTCTGCAG 99199
Qy 977 AAGCTGTTCCAGGATACATCAACCTCAACAGAACCTTCGGACTTTGCTGCTAAA 1036
Db 99200 AAGCTGATTCCAGGATACCATCGACCTCAACAGAACTCTTGGACTTTGCTGCTAAA 99259
Qy 1037 TTCTATGGACTGTACTGTGTGCAGGAGGTGGCAGAGACATTCGATTTGTGGTATGAAC 1096
Db 99260 TTCTATGGACTGTGTGTGTAAGGAGGTGGCAGAGACAGCAGATTTGCGGTGATGAAG 99319
Qy 1097 AATCTTTTACCAGATCGTAAAATGCATATCAAAATGACCTCAAGGCTCAACCTAC 1156
Db 99320 AATCTTTTACTAAGACTGCTCAAAATGCATACCAATATGACTTCAAAAGTTCAACCTAC 99379
Qy 1157 AAACGGCGGGTTCACGAAAGAGAGAGAGAGGCTC-----TCCACATTTAA 1206
Db 99380 AAAGAGCAGGCTTCACGAAAGAGAGAGAGAGGCTCTCCACATTTTCCACATTTAA 99439
Qy 1207 AGACCTAGACTCTTACAGACATCCCTGTGTTGTTTGTGATGCTGACATGACAA 1266
```







```
Qy 3112 ----- 3111
Db 101633 TTTCTCTTTTGAACAGTCTCACACTGTCGCTGGCTAGAGTGCAGTGGCGGATCTTG 101692
Qy 3112 ----- 3111
Db 101693 GCTCACTGCAACCTCCACCTCTCTGGTTCAAGCGATTCTCTGCTCAGCTCCCAAGTA 101752
Qy 3112 ----- 3111
Db 101753 GCTGGGATTACAGTGCCTGCCACCATGCCAGCTAATTTTGTGTTTTAGTAGAGAT 101812
Qy 3112 ----- 3111
Db 101813 GGGGTTTCACTATGTTAGCCAGGCTGGTCTCAACTCTCGACTCGTGATCCGCCACCTC 101872
Qy 3112 ----- 3111
Db 101873 GGCGCCCAAGTCTGGGTTACAGGCGTGAGCCACCGCGCCGCCACAGATATTTTC 101932
Qy 3120 TTAACATCATGGGAGACAGAGATCTTGTGCTGGTGAGTCAATGCTGTGCCATATGT 3179
Db 101933 TTAACATCATGGGAGACAGAGATCTTGTGCTCAGTGAAGTCAATGCTGTGCCATGT 101992
Qy 3180 CCTACCCCTCTCTTCATGACAGGAAGTTGGAATGGGGCTACATATGCCCTCTCCTC 3239
Db 101993 CCTA-CGCGCTGTCTTCATGACAGGAGGTTGGAATGGGGCGCCGATGTGCCCTCTCCT- 102050
Qy 3240 CCCGCTCTACAAGAGTTGTGTTTCCATCTGATCTCTCCACTCTGTGCAGGGGAAGAAGG 3299
Db 102051 CCCATCTACAAGAGTTGTGTTTCCATCTGATCTCTCCACTCTGTGTAGGGGAGGAGG 102110
Qy 3300 GGGCTGTGATCTCAGGAGATGTTGAATTCCTGTTCTATCCCTCTCTATCCACCT 3359
Db 102111 GGGCTGTGATCTCAGGAGATGTTGAATTCCTGTTCTATCCCTCTCTATCCCTCT- 102166
Qy 3360 GCCGTGATAATGTTAGCCATACCCCAATACTGCTATATAGACACCCCGAGCA 3419
Db 102167 ACCTTGATAGTAGTTAGCCATACCCCAATACTGCTATATAGACACCCCGAGCA 102226
Qy 3420 GTTCTGCTGCTGCTGCTCT- TTGCTGCCATGTTTTCACGAAGGAAG- AATCTTCTGCT 3477
Db 102227 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 102286
Qy 3478 ATTTTCTTTTTCATATTTTACATATTTATGATGATTTAATGCTGTTTATTAAGACAGACTT 3537
Db 102287 GCTATTTTTCATATTTTACATATTTATGATGATTTAATGCTGTTTATTAAGACAGACTT 102346
Qy 3538 CTGTTAGGGTGGGAGGAATATTGAGGAGGGCTGGGCTTAGGGAAGGAATGGGA 3597
Db 102347 CTGTAATGGTGGGAGGAATATTGAGGAGGGCTGGGCTTAGGGAAGGAATGGGA 102406
Qy 3598 AGCAACATTTTATTAAGTGTACTATTGCTCTCTACTTTCTGTTGTTGTCAGAAATGGCA 3657
Db 102407 ATCAACATTTATGAAGTGTCTTATTTGCTCTCTACTTTCTGTTGTTGTCAGAAATGGCA 102466
Qy 3658 ATACAATAAAGTATATATGCTTTTAAATGTAATAACATTTATGATGATTTTA 3713
Db 102467 ATGCAATAAAGTGTACATGTTTAAATGTAATAAATTTAATCATGATTTT 102522

RESULT 7
AF027390 225432 bp DNA linear PRI 02-APR-1998
LOCUS AF027390 Homo sapiens 7q telomere, complete sequence.
DEFINITION AF027390
ACCESSION AF027390.1 GI:3004858
VERSION AF027390.1
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS 1 (bases 1 to 225432)
Chi, H.-C., Saunders, E.H., Buckingham, J.M., Ricke, D.O., Munk, C.C.,
Lobb, R.R., Ueng, S.Y.-J., Mundt, M.O., White, P.S., Tatam, O.L.,
Riethman, H.C. and Moyzis, R.K.
TITLE DNA Sequence Analysis of the Terminal 236 kb of Human Chromosome 7q
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 225432)
Chi, H.-C., Saunders, E.H., Buckingham, J.M., Ricke, D.O., Munk, C.C.,
Lobb, R.R., Ueng, S.Y.-J., Mundt, M.O., Tatam, O.L., Riethman, H.C. and
Moyzis, R.K.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1997) Center for Human Genome Studies, Los Alamos
National Laboratory, MS M888, New Mexico, NM 87545, USA
FEATURES
source
1. .225432
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q telomere"
901..1496
/note="HSAL000641"
/rpt_family="Alu"
complement(2341..2376)
/note="GRAIL 2 good exon ; ORF; frame 0"
repeat_region
2513..2800
/note="M32220"
/rpt_family="retrovirus"
2659..2997
/note="LTR2"
4101..4283
/note="HSAL006653"
/rpt_family="Alu"
complement(4512..4600)
/note="MER3"
/rpt_family="MER3"
8171..8421
/note="GRAIL 2 good exon ; ORF; frame 0"
8929..9106
/note="GRAIL 2 marginal exon ; ORF; frame 1"
12304..12466
/note="Z69721"
/rpt_family="L1"
complement(12460..12687)
/note="Z69364"
/rpt_family="Alu"
complement(12621..12766)
/note="97% identity -- U66083"
complement(12792..14659)
/rpt_family="L1"
complement(14751..14845)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
19550..19636
/note="80% identity"
/db_xref="dbEST:R38951"
19655..19824
/note="80% identity -- Z77249"
19840..19998
/note="83% identity"
/db_xref="dbEST:W79023"
20128..20260
/note="81% identity -- Z77249"
21027..21097
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(24077..24350)
/note="U47924"
/rpt_family="Alu"
complement(24301..24492)
/note="HSAL000410"
/rpt_family="Alu"
24633..24901
/rpt_family="Alu"
complement(24992..25278)
/note="Z73359"
repeat_region
```

```

/rpt_family="Alu"
complement(25515. .25599)
/note="MALR00365"
/rpt_family="THE1"
complement(25777. .25901)
/note="MALR00322"
/rpt_family="MSTC"
complement(26808. .27050)
/note="X66298"
/rpt_family="Alu"
complement(27009. .27125)
/note="HSAL004729"
/rpt_family="Alu"
complement(27361. .27462)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(27482. .27721)
/note="GRAIL 2 good exon ; ORF; frame 2"
30922. 31155
/rpt_family="MER33"
31815. 31991
/note="GRAIL 2 good exon ; ORF; frame 1"
complement(331907. .34394)
/gene="VIP2R"
/product="vasoactive intestinal polypeptide receptor 2
precursor"
complement(31907. .32006)
/gene="VIP2R"
/note="99% identity -- L40764"
complement(join(<31908. .32006,34182. .34234))
/gene="VIP2R"
/codon_start=3
/product="vasoactive intestinal polypeptide receptor 2
precursor"
/db_xref="GI:3004859"
/protein_id="AAC09171.1"
/db_xref="GI:3004859"
/translation="MTLLPPLLCWLLAPVNSIHPCRFHLEIQEETKAEILLRS
QTEKHK"
complement(33411. .33597)
/gene="VIP2R"
/note="100% identity -- 259087"
complement(34182. .34232)
/gene="VIP2R"
/note="100% identity -- L40764"
complement(34200. .34394)
/gene="VIP2R"
/note="92% identity -- L36566"
34432. 34694
/rpt_family="TAR1"
35000. .35058
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(35325. .35445)
/rpt_family="MER42"
complement(36238. .36296)
/note="HSAL004342"
/rpt_family="Alu"
complement(37248. .37455)
/note="MER4_00001"
/rpt_family="MER4"
complement(37454. .37580)
/note="84% identity"
/db_xref="dbEST:T62489"
complement(37828. .38086)
/note="270040"
/rpt_family="MER4"
complement(41427. .41642)
/rpt_family="MLT2B2"
complement(42308. .42622)
/rpt_family="MLT2B2"
43022. 43522
/rpt_family="Tigger2"
43881. 44000

```

```

misc_feature
44984. .45116
/note="GRAIL 2 good exon ; ORF; frame 2"
misc_feature
complement(46422. .46478)
/note="GRAIL 2 good exon ; ORF; frame 0"
repeat_region
complement(46801. .47121)
/rpt_family="Alu"
complement(47511. .47779)
/note="GRAIL 2 good exon ; ORF; frame 1"
repeat_region
complement(48123. .48196)
/rpt_family="MLTc"
complement(49320. .49497)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
misc_feature
complement(49934. .50215)
/note="GRAIL 2 good exon ; ORF; frame 2"
misc_feature
complement(51833. .51901)
/note="GRAIL 2 good exon ; ORF; frame 2"
repeat_region
complement(52214. .52500)
/note="268162"
/rpt_family="Alu"
54960. 55032
/rpt_family="LIMC2"
55462. 55537
/rpt_family="LIME2"
57335. 58234
/rpt_family="LI"
complement(58350. .58466)
/rpt_family="MER44"
58503. 58646
/rpt_family="LIMA9"
complement(58757. .59020)
/note="GRAIL 2 good exon ; ORF; frame 0"
62145. 62345
/note="GRAIL 2 good exon ; ORF; frame 2"
repeat_region
complement(62518. .64744)
/rpt_family="Tigger1"
complement(66883. .67070)
/note="GRAIL 2 excellent exon ; ORF; frame 0"
misc_feature
complement(67293. .67336)
/note="GRAIL 2 good exon ; ORF; frame 1"
67421. 67746
/rpt_family="Alu"
complement(68255. .68320)
Query Match 55.1%; Score 2047.2; DB 9; Length 225432;
Best Local Similarity 75.0%; Pred. No 0;
Matches 3236; Conservative 0; Mismatches 403; Indels 677; Gaps 27;
QY 20 GAAGGACGAGAGGGCGTTCCTCTTTGGACATTTTCATCCCTCGTTTTTTTCAG 79
DB 120520 GGAGGACGAGCGGGCGTTCCTCGTGGGCTTCTCATGCCAGCTTTTCTCCAG 120579
QY 80 ATGTGGCTTGGTCTGGCGCAAGTCCAGCAGCCAGCTTAAGCTTACTTCTTGAAA 139
DB 120580 GTGTGACTTGGTC--GGACCGGATCCCGGCGCTAGTCTTCTGTGAAA 120637
QY 140 GGGGAAAGTATCCCTGTGGAAGCGTTAACTTTGGAGGGGTGGCGGACGTGAGTT 199
DB 120638 GAGTAAGTGCTCTTCGCGGAGAGCGGGTAACCTCCCAAGGCGGTTCGAGGCGGAGTT 120697
QY 200 CTTCCCCATGCCAGCGAATGGTGTGGCCTTGAGCTGGTCCAGGAGCGGCTCGACGTGT 259
DB 120698 C-TCCCCAGGCCAGCGAATGGGCGGCTGAGCTGGTCTA-GAGCGGCTCCACGCGT 120755
QY 260 CTGAGGAGCGCGGAGGGGGGGAGGTGGCCACACAGAACCGCGGTCTCTGAAGAGA 319
DB 120756 CTGAGGAGCGCGCGGGGCGGAGGAGGTGGCCAGAG--CGCGGGTCTCTGAAGAGA 120812
QY 320 CGTTGGGAAG---ATTTCGATTCCGAGAAGGAACCGGATTGAAGAGCCAGGCC 376
DB 120813 CTTGGGAGGAGAGATTCCATTGACGAGGAGGAGGAGGATTCAGAG--CCAGGCC 120871

```

Qy 377 GCTGAGGGGAGGGGCTGCTAAGATGCGCTCGGCTCTCCGGGCGCTGCTCTCGGTC 436  
|||||  
Db 120872 GCTGAGGGG--AGGGCTGCGCAAGATGGCCCTCGGTCTCTCAGGGCCCTCTCGGGGCTC 120929  
Qy 437 GGTTTTTCATCCTTTGATCCCGCGCTCCCTTCTCTGTACTTGTCTCTCAGCATCTGGAATC 496  
|||||  
Db 120930 AGTTTTTCATCCTTTGATCCAGGGGCTGCTTCTCTGTACTCTCTCAGCATCTGGAATC 120989  
Qy 497 AAGAGACCCATGGCATCTGAGGTGGCTTATGCTCTGCGATGCCATGCCATCAAGAAATAGGC 556  
|||||  
Db 120990 CAGAGACCTATGGGCTCTGAGGTGCTTGTGCTCGGCGATACCCATCAAGAAATAGGC 121049  
Qy 557 CATAGAAGTGTGATCTCTCAGGAGAGACAACATATAAAAGACAACCTCATCAGCCCTTG 616  
|||||  
Db 121050 CATGAGGTGTGATCTCTCAGGAGAGACAACCTATAGAAAGACGACCTCATCAGCCCTTG 121109  
Qy 617 AAAGGTGCCATCCAGTTAGGACTATCCACACACTGTGGGGAGCCCTGAGTACCAAAACAGAG 676  
|||||  
Db 121110 AAATGTGCCATCCAGTTGTGCATTACTCACACTGTGGGGAGCCCTGGATCCACACAGAG 121169  
Qy 677 CGTGATGTCCTCAGCAAGATTTCTACGTGGTTGAGAGTATCTTCTTTCCAGTGAAGGG 736  
|||||  
Db 121170 CGTCATGTCTCATGTAGATTTCTGCTGTGGGAGATATCTTCTTTCCAGTG- ---- 121224  
Qy 737 AGCAACCTGACCCCTGCTCATCAGTACATGACTTTCGTTTCAAGACCTATGCACTCTGT 796  
|||||  
Db 121225 -----GGACCCCTGCTCATCTGTCAGTGGAGTTTCTTTCAAGACCTATACACCTGTT 121277  
Qy 797 GCCTTCGCTACTTCTCGGGAGCTATTTGGTATCCGGCCGCGATGATTACTTCTTATTCCTC 856  
|||||  
Db 121278 GCCTTCGCTACTTTCAGGAGCTATTTGGTATCTGCCCTGATGATGACTTGTACTCCCTC 121337  
Qy 857 TGCAGTGAGCCGCTGATGTAAGTCTGTAGCTCTGAGGCTGATGGTTCCCTATTCATGTG 916  
|||||  
Db 121338 TGCAGTGAGCCGCTGATGTAAGTCTGTAGCTCTGAGGCTGATGGTTCCCTGTTACATGTG 121397  
Qy 917 TCCAGCGAGATGATGATTAAGACAGTCCCAACATAGAGCGGGAATTTCTGCAG 976  
|||||  
Db 121398 TCCATCCAGTGAATTCATTTAAACAGTCCCAACATAGAGCGGAGTTCTGCGAG 121457  
Qy 977 AAGCTGCTTCCAGGATACATGAACCTCAACAGAACCTCGGACTTTGCTGCTTAA 1036  
|||||  
Db 121458 AAGCTGATTCAGGATACCATCGACCTCAACAGAACTCTTGACTTTGCTGCTTAA 121517  
Qy 1037 TTCTATGAGCTGTACTGTGTCGAGGCTGGCAAGAACATTCGGATGTTGGTGATGAAC 1096  
|||||  
Db 121518 TTCTATGAGCTGTGCTGTGTGAAGGCAAGTGGCAAGAACAGCAGATTCGCGTGTGAAG 121577  
Qy 1097 AATCTTTTACCAGATCGGTAAATGCATATCAATATGACCTCAAGGCTCAACCTAC 1156  
|||||  
Db 121578 AATCTTTTACGACTGGTCAAAATGCATATCAATATGACCTCAAAAGTTCAACCTAC 121637  
Qy 1157 AAACGGGGGCTTCCAGAAAGCGAGAGAGGCTCTTCCACAC- -----TTTAA 1206  
|||||  
Db 121638 AAAGAGCGGCTTCCAGAAAGCAAGAGAGGCTCTTCCACATTTTCCACATTTAA 121697  
Qy 1207 AGACCTAGACTTCTTACAAGACATCCCTGTGATGGTCTTTTTTTGGATGCTGACATGACAA 1266  
|||||  
Db 121698 AAACCTAGACTTCTTACAAGACATCCCGATGGTCTTTTTTTGGATGCTGATGATGACAA 121757  
Qy 1267 CGCTCTCTGAAGACCTGCGAGCGTGAATGTTGCTGCTGAGAGCTTCAGATATGGA 1326  
|||||  
Db 121758 CGCTCTCTGAAGACCTGCGAGCGTGGCTGTTGGTGTGCGATGATCTTGGATATGGA 121817  
Qy 1327 TTACAGCTCTTGTATGTCAATCCATATATATAGATCATGCACAGAGAGCCCTTAAGAC 1386  
|||||  
Db 121818 CTATAGCTCTTGTATGTCAATCCACACATATATATG- ---ACGAGAGCCCTTGAGCAG 121874  
Qy 1387 TGAACACAGTACTCAGTTGATCTCGAAGACCGGCCCCCCCCAAAGGCTCTGTATTCAC 1446  
|||||  
Db 121875 CCAACACAAATGCTCAGTTCACACTTGAAGACCGGCCCCCCCCAAAGGCTCTGTATTCAC 121934

Qy 1447 AGCATGAATCATCCATCCAGGGAGAGGCTCGAGGGGTGTACCATGGAGACTGATGACCA 1506  
|||||  
Db 121935 AGCATGAATCATCCATCCAGGGAGAGGCTTGACGGGCGACCCCATGGAGACTGATGACCA 121994  
Qy 1507 TATGGGTGGCATCCCTGCCCCGAATAGTAAAGGGGAAAGGCTTCTGCTTTATATTTGGCAT 1566  
|||||  
Db 121995 TATGGGTGGCATCCCTGCCAGATAGTGAAGGGGAAAGAAATCCTTGTGTTATTTGGCAT 122054  
Qy 1567 CATTCACATCTACAGTCTTACAGTCTTACAGTCTTAAAGATTTGGAGACACTCTTTGAAAGCCCT 1626  
|||||  
Db 122055 CACTGACATCTTACAGTCTTACAGTCTTAAAGATTTGGAGCATTTCTTTGAAAGCCCT 122114  
Qy 1627 GGTACATCAGGAGACACTGTCTCAGTGCATCGCCAGGCTTCTAGCTGAAAGGTTCCA 1686  
|||||  
Db 122115 GGTATGTATGGGACACTGTCTCAGTGCATCGCCAGGCTTCTACACTGAACGGTTCCA 122174  
Qy 1687 GCGCTTCATGTGCAACACAGATTTTAAAGAAAT- ----- 1720  
|||||  
Db 122175 GTGCTTCATGTGCAACACAGCATTTAAGNAGATCCGGCCAGGTGCCGTGGCTCGCACCT 122234  
Qy 1721 ----- 1720  
Db 122235 GTAATCCAGGACTTTGGGAGGCCAAGCGGATGGATCACCTGAGTCTCAGGAGTTCAAGA 122294  
Qy 1721 ----- 1720  
Db 122295 CCAGTTGGCCAAACATGTTGTAACCCCATCTCTACTAAATCACAGAAATTAGCCGGCG 122354  
Qy 1721 ----- 1720  
Db 122355 GTAGTGGCACATGCTCTTAACCTCAGTACTTGGTTAGGCCAAGCAGGAGNATCACTGA 122414  
Qy 1721 ----- 1720  
Db 122415 GCCTAGGAGCAGAGGTTGTCAGTGAGCAAGATTTGCACCTGCACCTCCAGCTGGGCAA 122474  
Qy 1721 -----CCCTTGAAGCCTTCTCCTTCCAAA 1744  
Db 122475 AAGAGTGAGACCTGTCTCAAAAAAAGAAAGATTCCTTGAAGCCTTCTCCTTCCAAA 122534  
Qy 1745 AAGTTTCGCTGCTGCTCATCTTTCTCGGGCAGCAGGCTCCAGTGGCAACTCTCTGCAAT 1804  
|||||  
Db 122535 AAGTTTCGCTGCTGCTCATCTTTCTCTGGCGAGCAGGCTCCAGTGACAACTCTCTG- --- 122590  
Qy 1805 ACTTACAGCATCGCTCTCTGGGGAAACACAAGGCACAAGTGACACAACAAAGCGAAGTG 1864  
|||||  
Db 122591 TGTTACCAGCATCGCTCTCAGGGGAACACAAGGCACAACCAACAAAGCGGGAAGTG 122650  
Qy 1865 GAGCCAGCGGTTCACTTGGTCTCTGATGTTTACCTCAGACTCCACCTTTGGAGGAA 1924  
|||||  
Db 122651 GAGCCAGGTTTCACTTGGTGTCTCTGATGTTTACCTCAGACTCCACCTTTGGAGGAA 122710  
Qy 1925 ATCAGTGAGGCTCGCTTATTCCTGACCCAGTCTTCTCAGCTCTAGTTGGAGAGACTTTG 1984  
|||||  
Db 122711 ATCAGTGAGGTTTACCTTATTCCTGACCCAGTCTTCTCAGCTGATGTTGGAGAGACTTG 122770  
Qy 1985 CAATGCTTAACTACAAGTACAACCTTTGAAAAAGCTTGAAGTTGACAGTTCAGAGTTCACC 2044  
|||||  
Db 122771 CAATGCTTAACTACAAGTACAACCTTTGAAAAAGCTTGAATTTGACAGTTCAGAGTTCACC 122830  
Qy 2045 CATTAAGCGCAAGGCTCAGAGAGCTGGAACAAGATCTGCCATCTCTGTGATCCCAAG 2104  
|||||  
Db 122831 CATTAAGCGCAAGGCTCAGAGAGCTGGAACAAGATCTTGCATCTCTGTGATCCCAAG 122890  
Qy 2105 ATGTCAGCCCTTGGCCCGAGCAATGCTGAATTTCTTACTTGTGTTGTCATCAAAAAAGGACT 2164  
|||||  
Db 122891 ATGTCAGCCCTTGGCCCGAGCAATGCTGATTTCTTCTACTTGTGCCAT-TAAAAAGGAT 122949  
Qy 2165 GTAATGAAAGTGAAGGAGCTGCTCCTCCATCTTCTTCTTGAAGAAAGACCTTCTCTCT 2224  
|||||  
Db 122950 GTAATGAAAGGAGGAGGAGCT----CTCATCTTCTTCTTGAAGAAAGACCTTAGTCTCT 123005  
Qy 2225 TCCCTTCTCTATGATGAGGCTTAGTGCCTCAGAGAGTTGAGGACCGCAGCATCCCTC 2284

|||||  
Db 123006 TCCCCCTCTCATGAATGGGCATTTAGTGCTCAGACAGTTGAGGATGCGAGCAACCCCTC 123005  
QY 2285 CACTCCAGAGTTGGGTGGTACGGATTTTCAACTGGCCACCCCTTTGCCCTCCACTATTGAA 2344  
|||||  
Db 123066 CACTCCAGAATTGGGTGG-ATGGAATTTTCAATGGGCTAAC--TTGGCTTTCACAATTGA 123122  
QY 2345 TTTTTCAGACCCCATCTCTCATGCTGGAATGGGATTCGTGGACTTGGCAGCTTTCT 2404  
Db 123123 GCTTTTTCAGACCCCATCTCTCATGCTGGAATGGGATGGCTGGACTTGGCAGTTTCT 123182  
QY 2405 TTTCCCTCTCTTTGACTAGGAACCGGACTCTTAATTTCTCAGGACAGACTAGCTGGCA 2464  
|||||  
Db 123183 TTTCCCTCATCTTTCACAGGAGCTGGACTCAATAATTTCTCAGGACAGACTAGCTGGCA 123242  
QY 2465 CATATCCCTACCTATGTTCTTCTCTGACTCTCTGGAAGATACTCTCTGTAATCTCTG 2524  
Db 123243 CATATACCCACCTCAGTTCTTCTCTCTGATCCCTGGAAGAAGATCCCTGTAATCTCTG 123302  
QY 2525 TAAAGGTTTTGGGGGATGAAGGTGTTTAAACCACCTCCAGCTTCTCTCTCTTTTCTT 2584  
Db 123303 TAAAGGTTTTGGGGGATGAAGGTGTTTAAACCACCTCCCAACTTCTCTTTTATTTCT- 123361  
QY 2585 TTTCTGAAAAAGGAAAAAGCACACACACACAATTTCAAGCCATTTTCAGATCAGAACTC 2644  
Db 123362 ---GAATAAGGAAAAATGCGCACACACGCAATTTCAAGCCAGTTTCAGATCAAAACCC 123418  
QY 2645 CAGAAGTGTGCAAGATGCCATTTCTGTAGAGTTCCCTCAGAGAGCCATGTTGTTATG 2704  
Db 123419 CAAAAGTGTGAGGAGATGCCATGATCATAAGATTCCCTCAGAGAACCATTGGTGTG 123478  
QY 2705 AAGAGAAGAGTAGTATGCTGCTGCCAGAGCAGCTCTCTTTAACTCCTCTCTCTCTG 2764  
Db 123479 AAGAGAAGTGTGAGGATGCTCTGCCAGGA-CAGCTCTCTTTAACTCCTCTCTCTCTG 123537  
QY 2765 ATGAATTTCTTAAGGCTGAAGGAATCAAGAGAGTGGGACATGGGTAATCTTTATCCCTT 2824  
Db 123538 ATGAATTTGTTAAGGCTGAGGAAATGGAGAGTGGGACATGGGTAATCTTTACCCCTT 123597  
QY 2825 TTGTTAAACAGAGGAGCCATGAGGCTGGGAGATCATAGCCCTTCCAGGAGAACTCT 2884  
Db 123598 TTTTAAACATGGGCGAGCCGCTGGGCTGGGAGATCATAGCCCTTCCAGGAGAACTCT 123657  
QY 2885 GTTCACTGCCAGCTATAGTAATTTACTATTTTCAATTTTCAATATATCTGTTGT 2944  
Db 123658 ACTCACTGCCAGGCTAATGATTTACTGTTTTCGAATTTGAAATATATTC-GGTTGT 123716  
QY 2945 TTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATTTCCAGAGGAGATTTT 3004  
Db 123717 TTTTGTAAATATGAACCTATCAATGAATTTTAGATCATTTCCAGAGGAGATTTCTT 123776  
QY 3005 TTGCTCTCTCATCTTTTCCACAGAGTGTCTCTCTGTTTGTGGAGCTAAGGTAAGAGGG 3064  
Db 123777 TT-CTCTTCCCATCTTTTCCACAGTGTCTCTCTGTTTGTGGAGCTAAGGTA----GAG 123830  
QY 3065 ACATCTCTCTGTTTAAACAGACAGTCCATATCTGTGAGGCGAGCA-- 3111  
Db 123831 ACACCTCTCTGTTTAAACAGGAGTCCATATCTATGAGGCTAGCAAAATTTTCTTTTC 123890  
QY 3112 ----- 3111  
Db 123891 TTTTCTTTTGAACAGCTCTACACTGTCCCTGGCTAGAGTGCAAGTGGCGGATCTTG 123950  
QY 3112 ----- 3111  
Db 123951 GCTCACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCTGCTCAGCTCCCAAGTA 124010  
QY 3112 ----- 3111  
Db 124011 GCTGGATTACAGGTCCTGCCACCAAGGCCAGCTAAATTTTGTGTTTGTAGTAGAT 124070  
QY 3112 ----- 3111

Db 124071 GGGGTTTCACTATGTTAGCCAGGCTGGTCTCAACTCCTGACCTCGTGTGATCCGCCACCTC 124130  
QY 3112 -----ATATTTTC 3119  
Db 124131 GGCCCCCAAAGTGTGGGTTTACAGCGGTGAGCCACCGCGCCGCCACAGATATTTTC 124190  
QY 3120 TTAACATCATGGGACAGACAGATTTCTGCCCTTGTGTGAGTCAATTCGTGTGCCATATGT 3179  
Db 124191 TTAACATCATGAGGAGAGACAGATTTCTGCCCTCAGTCAAGTCAATTCGTGTGCCATATGT 124250  
QY 3180 CCTTACCCCTCTGTTCTATGTCAGGGAAGTTGGAATGGGGCTTACATATATGCCCTCTCCCT 3239  
Db 124251 CCTA-CCCGCTGTTCTATGTCAGGAGGTTGGAATGGGGCCGATGTGCCCTCTCCT- 124308  
QY 3240 CCGCTTACAAGAGTTGTGGTTTTCCATCTGATCCTTCCACTCTTGTGAGGGGAAGAAG 3299  
Db 124309 CCCATCTACAAGAGTTGTGGTTTTCCCATCTGATCCTTCCACTCTTGTAGGGGAGGAAG 124368  
QY 3300 GGGCCTGGTATCTCAGGACAGATTTGTAATCCTGTTCTATCCTCTCTATCCCACTT 3359  
Db 124369 GGGCCTGGTATCTCAGGACAGATTTGTAATCCTGTTCTATCCTCTTCTCCTCCT- 124424  
QY 3360 GCCTTCAATATATGTTAGCCCATACCCCAATAAATCTCTATATTAGACACCCCGAGCCA 3419  
Db 124425 ACCTTGATAGTAGTTAGCCCATACCCCAATAAATCTCTATATTAGACACCCCGAGCCA 124484  
QY 3420 GTTTCGGCTGCTGTCT-TTGCTGCTGCTGTTTTCACAGAGGAAG-ATATCTTGTCT 3477  
Db 124485 GTTTCGGCTGCTGTCTGTCATGTTCTTTTACAAGAGGAAGAAATAT 124544  
QY 3478 ATTTTCTTTCATAATTTACTATTTATGATGTAATTAAGTGTGTTTATTAAGGACAGATT 3537  
Db 124545 GCTATTTTTCATAATTTACTATTTATGAAGTATTTAAGTGTGTTTATTAAGTATAGATT 124604  
QY 3538 CTGTAGGGGTGGGAGGAATATTTGAGGAGGCGTGGTCTTTAGGAAAGGAATGGGA 3597  
Db 124605 CTGTATGGGTGGGAGGAATATTTGAGGAGGCGTGGTCTTTAAGAAAGGAATGGGA 124664  
QY 3598 AGCAACATTTTATTAAGTGTACTATTTGCCCTCTACTTTGTATTTTTCAGAAATGGCAA 3657  
Db 124665 ATCAACATTTATTAAGTGTCTCTATTTGCCCTCTACTTTGTATTTTCAGAAATGGCAA 124724  
QY 3658 ATCAATATAAAGTCATATATGTTTAAATGTAATAAATTTAATGAGTTATTTA 3713  
Db 124725 ATGCAATATAAAGTGTACATGTTTAAATGTAATAAATTTAATGAGTTTATTA 124780

RESULT 8  
HSU78576  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
Source

HSU78576 2133 bp mRNA linear PRI 20-DEC-1996  
Human 68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha  
mRNA, clone PIP5K1a2, complete cds.  
U78576  
U78576.1 GI:1743872  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Loijens, J.C. and Anderson, R.A.  
Type I phosphatidylinositol-4-phosphate 5-kinases are distinct  
members of this novel lipid kinase family  
J. Biol. Chem. 271 (51), 32937-32943 (1996)  
97115834  
8955136  
2 (bases 1 to 2133)  
Loijens, J.C. and Anderson, R.A.  
Direct Submission  
Submitted (18-NOV-1996) Pharmacology, University of Wisconsin -  
Madison, 1300 University Ave., Madison, WI 53706, USA  
Location/Qualifiers  
1. .2133

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone_lib="pip5K1a2"
/clonetechnology="human fetal brain 5'-stretch cDNA library"
/clone_id="HL1149x"
/EC_number="2.7.1.68"
/notes="isoforms, possibly alternatively spliced, encoded by GenBank Accession Numbers U78575 and U78577"
/codon_start=1
/product="68 kDa type I phosphatidylinositol-4-phosphate 5-kinase alpha"
/protein_id="AAC50911.1"
/db_xref="GI:1743873"
/translation="MASASSGSSSVFSSDFPAVPSCITLSSASGKIKPMASEVLEA
KQDYLIVPYAGSMPIKIKHRSVDSGGTTTKKTKSSALKAGIQGITHTVGSLSST
KPRDVLMDYFVVEIFFPSEGSNTPAHYNDFRFKYAPVAFYFRELFGIRPDND
YLYSLCSPELIECSGAGSLFVSSDDDEFIKTKVQHEAEFLQKLLPQYWNKLNQD
PRLFLKFTGLYCVQAGKNIRIVWNNLLPRSVKHIKVDLKGSTYKRRASQREK
PLPFLDLPQIDPDLADMYNALCKTLQDCLVLQSEKIMDYSLIMS1HNIDH
AOREPLSSTQYSDVTRRPAPOKALYSTAMESIQGEARRGTMETDDHMGKIPARNK
GERLLLYIGIIDLQYRFVKLEHSHKALVHOGDVSVHRKPGYFAERFORFQNTVF
KKIPLKPSKPKRSFSSFRNRAGSSNSCITYOPSVSGEKKAOVTKAEVPEGVHLG
RPDLVLPQTPLEIESGSPIDPFSFPLVGETLQMLTSTTLEKLEVAESEFTH"
2110..>2133
3'UTR
BASE COUNT 524 a 524 c 559 g 526 t
ORIGIN

```

```

Query Match 54.2%; Score 2014; DB 9; Length 2133;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 2073; Conservative 0; Mismatches 0; Indels 39; Gaps 2;

QY 1 ATTAACAGCGCGTGTAGGAAGACGAGAGAGGGCGTTCGCTTGGGACCTTTTCA 60
DB 22 ATTAACAGCGCGTGTAGGAAGACGAGAGAGGGCGTTCGCTTGGGACCTTTTCA 81
QY 61 TGCCTCGCTTTTTCAGATGTGGCTTGGTCTGGGCGCAAGGTCCAGAGCCAGCTTA 120
DB 82 TGCCTCGCTTTTTCAGATGTGGCTTGGTCTGGGCGCAAGGTCCAGAGCCAGCTTA 141
QY 121 AGCTTACTCTTGTGAAGAGGAAAGTATCCCTGTGGAAAGCGGTAACTTGTGGAG 180
DB 142 AGCTTACTCTTGTGAAGAGGAAAGTATCCCTGTGGAAAGCGGTAACTTGTGGAG 201
QY 181 GGGTGGCGGACGTGAGTCTTCCCATGCCAGCGAATGGTGGCCCTGAGCTGGTCC 240
DB 202 GGGTGGCGGACGTGAGTCTTCCCATGCCAGCGAATGGTGGCCCTGAGCTGGTCC 261
QY 241 AGGAGCGCGCTCGACGTGTCTGAGGAGGCGCGGAGGGCGGAGGTGGCCACACGAA 300
DB 262 AGGAGCGCGCTCGACGTGTCTGAGGAGGCGCGGAGGGCGGAGGTGGCCACACGAA 321
QY 301 CGCGGGTCTGTAAAGAGAGCTGGGAGAGATTCGATTCGAGAGAGAGAGAACCGGATT 360
DB 322 CGCGGGTCTGTAAAGAGAGCTGGGAGAGATTCGATTCGAGAGAGAGAGAACCGGATT 381
QY 361 GAAAGAGAGCCAGCCGCTGAGGGGAGGGGCTGCTAAGATGCGTGGCTCCTCCGG 420
DB 382 GAAAGAGAGCCAGCCGCTGAGGGGAGGGGCTGCTAAGATGCGTGGCTCCTCCGG 441
QY 421 GCCGTCTCTTGGTGGGTTTTCATCCTTTGATCCGCGGCTCCTTCCCTGATCTTGC 480
DB 442 GCCGTCTCTTGGTGGGTTTTCATCCTTTGATCCGCGGCTCCTTCCCTGATCTTGC 501
QY 481 CT---CAGCATCTGAATCAAGACACCCATGGCATCTGA----- 516
DB 502 CTCAGCAGCATCTGAATCAAGACACCCATGGCATCTGA----- 561
QY 517 -----GCTGCTTATGCTCTGGCATCCCATCAAGAAATAGGCCATAG 561
DB 562 TTCTTACATCTCATTTGTGGCTTATGCTCTGGCATCCCATCAAGAAATAGGCCATAG 621
QY 562 AAGTGTGATTCTCTCAGGAGAGACAACATATAAAGACACACCTCATCAGCCTTGAAGG 621

```

```

DB 622 AAGTGTGATTCTCTCAGGAGAGACAACATATAAAGACACCTCATCAGCCTTGAAGG 681
QY 622 TGCATCCAGTTAGGCAATACCCACACTGTGGGAGCCTGAGTACCAACACGAGCGTGA 681
DB 682 TGCATCCAGTTAGGCAATACCCACACTGTGGGAGCCTGAGTACCAACACGAGCGTGA 741
QY 682 TGTCTCATGCAAGATTCTACGTTGGTGTGAGAGTATCTTCTTCCAGTGAAGGAGCAA 741
DB 742 TGTCTCATGCAAGATTCTACGTTGGTGTGAGAGTATCTTCTTCCAGTGAAGGAGCAA 801
QY 742 CTGACCCCTGCTCATCACTACTCAATGACTTTCGTTTCAAGACCTATGACCTGTTCCTT 801
DB 802 CTGACCCCTGCTCATCACTACTCAATGACTTTCGTTTCAAGACCTATGACCTGTTCCTT 861
QY 802 CCGTACTTCCGGGAGCTATTTCGTTATCCGCGCCGATGATTACTTGTATTCCTCTGCG 861
DB 862 CCGTACTTCCGGGAGCTATTTCGTTATCCGCGCCGATGATTACTTGTATTCCTCTGCG 921
QY 862 TGAGCCGCTGATTGACTCTGTAGCTCTGGAGCTAGTGGTCCCTTATCTATGTGTCCAG 921
DB 922 TGAGCCGCTGATTGACTCTGTAGCTCTGGAGCTAGTGGTCCCTTATCTATGTGTCCAG 981
QY 922 CGAGATGAGTTCATTATTAAGACAGTCCCAATATAAGAGCGGGAATTTCTGCAAGACT 981
DB 982 CGAGATGAGTTCATTATTAAGACAGTCCCAATATAAGAGCGGGAATTTCTGCAAGACT 1041
QY 982 GCTTCCAGGATPACTACATGAACCTCAACAGAACCCCTCGGACTTTGCTGCCTTAATTTCTA 1041
DB 1042 GCTTCCAGGATPACTACATGAACCTCAACAGAACCCCTCGGACTTTGCTGCCTTAATTTCTA 1101
QY 1042 TGGACTGTACTGTGTGAGGAGCTGGCAAGACATTTCGGATTGTGGTGTATGAACAATCT 1101
DB 1102 TGGACTGTACTGTGTGAGGAGCTGGCAAGACATTTCGGATTGTGGTGTATGAACAATCT 1161
QY 1102 TTTACCAAGATCGGTAAATATGCATATCAATATGACCTCAAGGCTCAACCTCAACAG 1161
DB 1162 TTTACCAAGATCGGTAAATATGCATATCAATATGACCTCAAGGCTCAACCTCAACAG 1221
QY 1162 CGGGCTTCCAGAAAGAGCGAGAGAGCCCTTCTCCACATTTAAAGACCTAGACTTCTT 1221
DB 1222 CGGGCTTCCAGAAAGAGCGAGAGAGCCCTTCTCCACATTTAAAGACCTAGACTTCTT 1281
QY 1222 ACAAGACATCCCTGATGCTCTTTTGGATGCTGACATGTACACGCTCTCTGTAAGAC 1281
DB 1282 ACAAGACATCCCTGATGCTCTTTTGGATGCTGACATGTACACGCTCTCTGTAAGAC 1341
QY 1282 CTTGACGCTGACTGTTTGGTGTGAGAGCTTCAAGATAATGGATTACAGCTCTTGTAT 1341
DB 1342 CTTGACGCTGACTGTTTGGTGTGAGAGCTTCAAGATAATGGATTACAGCTCTTGTAT 1401
QY 1342 GTCAATCCATATATAGATCATGCACACGAGAGCCCTTAAAGCAGTGAACACAGTACTC 1401
DB 1402 GTCAATCCATATATAGATCATGCACACGAGAGCCCTTAAAGCAGTGAACACAGTACTC 1461
QY 1402 AGTTGATCTCGAGAGCGGCGCCCAAGAGCTCTGTATTCACAGCCATGGAATCCAT 1461
DB 1462 AGTTGATCTCGAGAGCGGCGCCCAAGAGCTCTGTATTCACAGCCATGGAATCCAT 1521
QY 1462 CCAGGAGAGGCTCGAGCGGCTGTACCATGGAGACTCATGACCATATGGGTGGCATCCC 1521
DB 1522 CCAGGAGAGGCTCGAGCGGCTGTACCATGGAGACTCATGACCATATGGGTGGCATCCC 1581
QY 1522 TGCCCGGAATAGTAAAGGGAAAGGCTTCTGCTTTATATTGGCATCATTTGACATTTTACA 1581
DB 1582 TGCCCGGAATAGTAAAGGGAAAGGCTTCTGCTTTATATTGGCATCATTTGACATTTTACA 1641
QY 1582 GTCCTACAGGTTTGTAAAGAGTGGAGCACTCTTGGAAAGCCCTGGTGTACATGACGGAGA 1641
DB 1642 GTCCTACAGGTTTGTAAAGAGTGGAGCACTCTTGGAAAGCCCTGGTGTACATGACGGAGA 1701
QY 1642 CACTGTCTCAGTGCATGCCAGGCTTCTAGCTGAACGCTTCCAGCCTTCATCTGCA 1701

```



```
Db 842 TTACTTGTATTCCTCTCCAGTACGCGCTGATTGAACCTCTGTAGCTCTCTGGAGCTAGTGG 901
Qy 901 TTCCCTATCTATCTGTCCAGCGAGTATGAGTTTCAATATTAAGACAGTCCACATTAAGA 960
Db 902 TTCCCTATCTATCTGTCCAGCGAGTATGAGTTTCAATATTAAGACAGTCCACATTAAGA 961
Qy 961 GCGGAAATTTCTGCAGAGCTGCTTCCAGGATCTATCATGAACCTCAACAGAACCCCTCG 1020
Db 962 GCGGAAATTTCTGCAGAGCTGCTTCCAGGATCTATCATGAACCTCAACAGAACCCCTCG 1021
Qy 1021 GACTTGTGCTGCTAAATCTATGAGCTGTACTGTGTGAGCGAGGTGGCGAAGACATTCG 1080
Db 1022 GACTTGTGCTGCTAAATCTATGAGCTGTACTGTGTGAGCGAGGTGGCGAAGACATTCG 1081
Qy 1081 GATTGTGTGATGATCAACATCTTTTACCAAGATCGTAAATGCATATCAATATGACCT 1140
Db 1082 GATTGTGTGATGATCAACATCTTTTACCAAGATCGTAAATGCATATCAATATGACCT 1141
Qy 1141 CAAAGGCTCAACCTTACAAACGCGGGCTTCCAGAAAGAGGAGAGAGCCCTCTTCCAC 1200
Db 1142 CAAAGGCTCAACCTTACAAACGCGGGCTTCCAGAAAGAGGAGAGAGCCCTCTTCCAC 1201
Qy 1201 ATTTAAGACCTAGACTCTTTACAGACATCCCTGATGGTCTTTTGGATGCTGACAT 1260
Db 1202 ATTTAAGACCTAGACTCTTTACAGACATCCCTGATGGTCTTTTGGATGCTGACAT 1261
Qy 1261 GTACAAAGCTCTCTGTAAAGCCCTGCAGCGTACTGTGTGCTGAGAGCTTCAAGAT 1320
Db 1262 GTACAAAGCTCTCTGTAAAGCCCTGCAGCGTACTGTGTGCTGAGAGCTTCAAGAT 1321
Qy 1321 AATGATTACAGCTCTGTATGATCAATCCATATATAGATCATGCACACAGAGCCCTT 1380
Db 1322 AATGATTACAGCTCTGTATGATCAATCCATATATAGATCATGCACACAGAGCCCTT 1381
Qy 1381 AAGCAGTGAACACAGTACTAGTTGATCTGCAAGACCGGCCCCCAAAAGGCTCTGTA 1440
Db 1382 AAGCAGTGAACACAGTACTAGTTGATCTGCAAGACCGGCCCCCAAAAGGCTCTGTA 1441
Qy 1441 TTCACAGCCATGAATCCATCCAGGAGAGGCTCGAGGGTGGTACCATGGAGACTGA 1500
Db 1442 TTCACAGCCATGAATCCATCCAGGAGAGGCTCGAGGGTGGTACCATGGAGACTGA 1501
Qy 1501 TGACATATGGTGGCTGCTCCGCGGAAAGTAAAGGAAAGGCTTCGCTTTATAT 1560
Db 1502 TGACATATGGTGGCTGCTCCGCGGAAAGTAAAGGAAAGGCTTCGCTTTATAT 1561
Qy 1561 TGGCATCATTTGACATTTACAGTCTTACAGTTTGTAAAGTGGAGCACTCTTGGAA 1620
Db 1562 TGGCATCATTTGACATTTACAGTCTTACAGTTTGTAAAGTGGAGCACTCTTGGAA 1621
Qy 1621 AGCCCTGGTACATCAGGAGACATCTGTCTAGTGCATCGCCAGGCTTCTAGCTGAACG 1680
Db 1622 AGCCCTGGTACATCAGGAGACATCTGTCTAGTGCATCGCCAGGCTTCTAGCTGAACG 1681
Qy 1681 GTTCCAGGCTTCATGTGCAACAGATTTTAAAGATTCCTTGAAGCCTTCTCCTTC 1740
Db 1682 GTTCCAGGCTTCATGTGCAACAGATTTTAAAGATTCCTTGAAGATTCCTTCCT 1725
Qy 1741 CAAAAAGTTTGGGTCTGGCTCATCTTCTCTCGGCGAGAGGCTCCAGTGGCAACTCCTG 1800
Db 1726 ----- 1725
Qy 1801 CATTACTTACCAGCCATCGGTCTCTGGGAAACAAGGACAAAGTGAACAAAGGACAGA 1860
Db 1726 ----- 1725
Qy 1861 AGTGGAGCCAGGCTTCACTTGTGTCTGTCTGTATGTTTACCTCAGACTCCACCTTTGGA 1920
Db 1726 -----GCGTTCACTTGTGTCTGTCTGTATGTTTACCTCAGACTCCACCTTTGGA 1774
Qy 1921 GGAATCAGTGAGGCTCGCTATCTCTGACCCAGTTTCTCACTCTAGTTGGAGAGAC 1980
Db 1775 GGAATCAGTGAGGCTCGCTATCTCTGACCCAGTTTCTCACTCTAGTTGGAGAGAC 1834
```

```
Qy 1981 TTTGCAATGCTAACTACAAGTACAACTTGAAAGGCTTGAAGTTGCAGAGTACAGATT 2040
Db 1835 TTTGCAATGCTAACTACAAGTACAACTTGAAAGGCTTGAAGTTGCAGAGTACAGATT 1894
Qy 2041 CACCCATTAAAGCGGAAAGGCTTGAAGGACCTTGAAGGATTTCTGCCATCTCTGTGATCC 2100
Db 1895 CACCCATTAAAGCGGAAAGGCTTGAAGGACCTTGAAGGATTTCTGCCATCTCTGTGATCC 1954
Qy 2101 CAAGATGTCAGCCCTTGCCCGAGCAATGCTGAATTTCTTACTTTGGTCAATCAA 2156
Db 1955 CAAGATGTCAGCCCTTGCCCGAGCAATGCTGAATTTCTTACTTTGGTCAATCAA 2010
```

RESULT 10  
AL365510/c

## LOCUS

DEFINITION Human DNA sequence from clone RP11-429H9 on chromosome 10, complete sequence.

## ACCESSION

AL365510

## VERSION

AL365510.32 GI:14348465

## KEYWORDS

HTG.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 180189)

## AUTHORS

Bray-Allen, S.

## TITLE

Direct Submission

## JOURNAL

Submitted (08-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

## COMMENT

requests: clonerequest@sanger.ac.uk  
On Jun 11, 2001 this sequence version replaced gi:14272312.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sv., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
RP11-429H9 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6  
This sequence is the entire insert of clone RP11-429H9. The true left end of clone RP11-162K11 is at 109726 in this sequence. The true right end of clone RP11-397J18 is at 42873 in this sequence.  
Location/Qualifiers  
1..180189  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-429H9"  
/clone\_lib="RPCI-11.2"  
1..224  
repeat\_region  
/note="MLT2FB repeat: matches 1..227 of consensus" 236..382  
repeat\_region  
/note="L1MC4 repeat: matches 7618..7776 of consensus"

```

misc_feature      complement(495..1074)
repeat_region    /note="match: STS: Em:HS100C7"
repeat_region    2802..3002
repeat_region    /note="MIR repeat: matches 46..244 of consensus"
repeat_region    4688..5178
repeat_region    /note="L1P13 repeat: matches 5648..6144 of consensus"
repeat_region    5188..5501
repeat_region    /note="AluX repeat: matches 1..312 of consensus"
repeat_region    5959..6225
repeat_region    /note="AluX repeat: matches 2..284 of consensus"
repeat_region    6739..6839
repeat_region    /note="MIR repeat: matches 80..191 of consensus"
repeat_region    6905..7200
repeat_region    /note="AluB repeat: matches 3..299 of consensus"
repeat_region    7268..8052
repeat_region    /note="L1A10 repeat: matches 5502..6287 of consensus"
repeat_region    8319..8491
repeat_region    /note="FRAM repeat: matches 0..173 of consensus"
repeat_region    9136..9446
repeat_region    /note="AluY repeat: matches 1..311 of consensus"
repeat_region    9447..9731
repeat_region    /note="AluJ repeat: matches 1..291 of consensus"
repeat_region    10082..10259
repeat_region    /note="FRAM repeat: matches -3..169 of consensus"
repeat_region    10265..10585
repeat_region    /note="AluX repeat: matches 1..301 of consensus"
repeat_region    10667..10910
repeat_region    /note="MIR repeat: matches 1..241 of consensus"
repeat_region    12586..12885
repeat_region    /note="AluS repeat: matches 1..300 of consensus"
repeat_region    13009..19121
repeat_region    /note="L1P15 repeat: matches 11..6144 of consensus"
repeat_region    19127..19218
repeat_region    /note="46 copies 2 mer aa 63% conserved"
repeat_region    19221..19287
repeat_region    /note="L2 repeat: matches 2574..2634 of consensus"
repeat_region    19341..19579
repeat_region    /note="L1B18 repeat: matches 5400..5638 of consensus"
repeat_region    19581..19923
repeat_region    /note="MER61A repeat: matches 5..354 of consensus"
repeat_region    19925..20445
repeat_region    /note="L1B18 repeat: matches 5640..6171 of consensus"
repeat_region    20641..20958
repeat_region    /note="AluX repeat: matches 1..298 of consensus"
repeat_region    21217..21766
repeat_region    /note="L1M4 repeat: matches 6342..6929 of consensus"
repeat_region    21772..21985
repeat_region    /note="AluB repeat: matches 86..297 of consensus"
repeat_region    21992..22122
repeat_region    /note="L1M4 repeat: matches 6934..7054 of consensus"
repeat_region    22123..22418
repeat_region    /note="AluX repeat: matches 1..301 of consensus"
repeat_region    22419..23122
repeat_region    /note="L1M4 repeat: matches 7054..7829 of consensus"
repeat_region    23182..23884
repeat_region    /note="L1M4 repeat: matches 6955..7618 of consensus"
repeat_region    23885..24166
repeat_region    /note="AluB repeat: matches 1..269 of consensus"
repeat_region    24167..24275
repeat_region    /note="L1M4 repeat: matches 7618..7730 of consensus"
repeat_region    24451..24542
repeat_region    /note="L2 repeat: matches 2624..2709 of consensus"
repeat_region    24755..25030
repeat_region    /note="MLT1A2 repeat: matches 6..313 of consensus"
repeat_region    25086..25135
repeat_region    /note="25 copies 2 mer tc 80% conserved"
repeat_region    25149..25458
repeat_region    /note="AluJ repeat: matches 1..296 of consensus"
repeat_region    25462..25535
repeat_region    /note="MLT1A2 repeat: matches 321..394 of consensus"
repeat_region    25943..26296
repeat_region    /note="MLT1A1 repeat: matches 1..343 of consensus"
repeat_region    29101..29479

/note="L2 repeat: matches 2156..2525 of consensus"
29481..30004
/note="L1B18 repeat: matches 5631..6153 of consensus"
30007..30210
/note="L2 repeat: matches 2531..2750 of consensus"
30338..30439
/note="MIR repeat: matches 57..165 of consensus"
30518..30576
/note="L2 repeat: matches 2690..2750 of consensus"
31508..31562
/note="L2 repeat: matches 2646..2704 of consensus"
31514..31717
/note="MIR repeat: matches 53..256 of consensus"
33399..33435
/note="tRNA-Leu-CTT repeat: matches 1..37 of consensus"
33482..33521
/note="MER5B repeat: matches 138..178 of consensus"
33523..33653
/note="FLAM_A repeat: matches 13..142 of consensus"
33673..33799
/note="MER5B repeat: matches 1..129 of consensus"
34148..35176
/note="L1B2 repeat: matches 4995..6089 of consensus"
35569..35673
/note="L2 repeat: matches 2634..2741 of consensus"
36553..36723
/note="AluSg/x repeat: matches 121..291 of consensus"
37781..38293
/note="MER21B repeat: matches 5..498 of consensus"
38317..38680
/note="L1P2 repeat: matches 5782..6146 of consensus"
38795..38944
/note="MER77 repeat: matches 462..791 of consensus"
39148..39249
/note="Tigger3b repeat: matches 1..102 of consensus"
39248..39571
/note="Tigger3b repeat: matches 274..584 of consensus"
39572..39928
/note="THE1B repeat: matches 1..364 of consensus"
39929..40526
/note="Tigger3b repeat: matches 584..1231 of consensus"
40527..40662
/note="MIR repeat: matches 16..154 of consensus"
41996..42091
/note="48 copies 2 mer tt 70% conserved"
42094..42372
/note="AluB repeat: matches 6..310 of consensus"
43374..43488
/note="MIR repeat: matches 138..262 of consensus"
43634..43785
/note="MER5B repeat: matches 2..156 of consensus"
43800..43895
/note="MER5B repeat: matches 2..130 of consensus"
43896..43979
/note="42 copies 2 mer tt 78% conserved"
complement(44191..44387)
/note="match: STS: Em:G34800"
47248..47367
/note="MER5A repeat: matches 50..169 of consensus"
47252..47394
/note="MER5A repeat: matches 36..183 of consensus"
47445..47516
/note="MSTB repeat: matches 1..73 of consensus"
47552..47619
/note="L2 repeat: matches 2641..2709 of consensus"
47628..47683
/note="MLT2B repeat: matches 394..448 of consensus"
47684..47815
/note="66 copies 2 mer ta 61% conserved"
47842..48211

```

Query Match  
Best Local Similarity

49.4% Score 1835; DB 9; Length 180189;  
95.9% Pred. No. 0;



Matches 1921: Conservative 0; Mismatches 65; Indels 17; Gaps 3;

Qy 1 ATTTAACAGCGCGTGGTTAGGAAGACGAGGAGAGGGGCGTCCCTTTGGGACTTTTCA 60  
 Db 47206 ATTTAACAGCGCGTGGTTAGGAAGACGAGGAGAGGGGCGTCCCTTTGGGACTTTTCA 47147  
 Qy 61 TGCCTCG-TTTTTCAGATGTGGTCTGTGGCGCAAGGTCCAGCAGCCAGCTT 119  
 Db 47146 TGCCTCGTTTTTTCAGATGTGGTCTGTGGCGCAAGGTCCAGCAGCCAGCTT 47087  
 Qy 120 AAGCTTACTCTCTGTGAAAGGGGAAAGTATCCCTCTGTGAAAGCGGTAAACTTTGGA 179  
 Db 47086 AAGCTTACTCTCTGTGAAAGGGGAAAGTATCCCTCTGTGAAAGCGGTAAACTTTGGA 47027  
 Qy 180 GGGGGTGGGACGTGAGTCTTCCCATGCCAGCGCAATGGTGTGGCTTTAGCTGGTC 239  
 Db 47026 GGGGGTGGGACGTGAGTCTTCCCATGCCAGCGCAATGGTGTGGCTTTAGCTGGTC 46967  
 Qy 240 CAGGAGCGGCTCGACGTGTCTGAGGAGGCGCGAGGGGCGGGAGGTGGCCACAGA 299  
 Db 46966 CAGGAGCGGCTCGACGTGTCTGAGGAGGCGCGAGGGGCGGGAGGTGGCCACAGA 46907  
 Qy 300 ACGCGGGTCTGTAAAGAGAGCTTTGGGAAGATTCGATTCCGAGAGAGGAACCGGAT 359  
 Db 46906 ACGTGGGTCTGTAAAGAGAGCTTTGGGAAGATTCGATTCCGAGAGAGGAACCGGAT 46847  
 Qy 360 TGAAGAGAGCGCGCTCAGGGGAGGGGCTGTAAAGTGGCTGGCCCTCTCCG 419  
 Db 46846 TGAAGAGAGCGCGCTTTCAGGGGAGGGGCTGTAAAGTGGCTGGCCCTCTCCG 46787  
 Qy 420 GCGCGTCTCTCGGTGGTTTTTCATCTTTGTATCCCGGCTCCCTTCCTGTACCTGT 479  
 Db 46786 GCGCGTCTCTCGGTGGTTTTTC -TCCGTTGATCCCGGCTCCCTTCCTGTACCTGT 46728  
 Qy 480 CCTCAGCATCTGGAATCAAGAGACCCATGGCATCTGAGGTGCCCTTATGCCCTCTGGCATGC 539  
 Db 46727 CCTCAGCATCTGGAATCAAGAGTCCCATGGCATCTGAGGTGCCCTTATGCCCTCTGGCATGC 46668  
 Qy 540 CCATCAAGAAAATAGGCATAGAAGTGTGATTCCTCAGGAGAGACAACATATAAAGA 599  
 Db 46667 CCATCAAGAAAATAGGCATAGAAGTGTGATTCCTCAGGAG-----GGA 46623  
 Qy 600 CAACCTCATAGCCTTGAAGGTGCCATCCAGTTAGGCATTACCCACACTGTGGGAGCC 659  
 Db 46622 CAACCTCATAGCCTTGAAGGTGCCATCCAGTTAGGCATTACCCACACTGTGGGAGCC 46563  
 Qy 660 TGAGTACCAACACAGAGCGTGTCTCTCATGCAAGATTTCTAGCTGGTTGAGAGTATCT 719  
 Db 46562 TGAGTACCAACACAGAGTGTCTCTCATGCAAGATTTCTAGCTGGTTGAGAGTATCT 46503  
 Qy 720 TCTTTCCAGTGAAGGAGCAACCTGACCCCTGTCTCATCAATGACTTTTGGTTTCA 779  
 Db 46502 TCTTTCCAGTGAAGGAGCAACCTGACCCCTGTCTCATCAATGACTTTTGGTTTCA 46443  
 Qy 780 AGACTATGACCTTGTTCCTTCGCTTCCGCTACTTTCGCGAGCTATTGGTATCCGCGCCGATG 839  
 Db 46442 AGACTATGACCTTGTTCCTTCGCTTCCGCTACTTTCGCGAGCTATTGGTATCCGCGCCGATG 46383  
 Qy 840 ATTACTTGTATTCCTCTGCAAGTGGCGCTGATTGAACTCTGTAGCTCTGGAGCTAGT 899  
 Db 46382 ATTACTTGTATTCCTCTGCAAGTGGCGCTGATTGAACTCTGTAGCTCTGGAGCTAGT 46323  
 Qy 900 GTTCCCTATCTATGTCTCAGCAGCAAGTGTGATTTTAAAGACAGTCCCAATAAAG 959  
 Db 46322 GTTCCCTATCTATGTCTCAGCAGCAAGTGTGATTTTAAAGACAGTCCCGACATAAAG 46263  
 Qy 960 AGGCGGAATTTCTCAGAGCTGCTCCAGGATACATCACTCAACCTCAACGACCCCTC 1019  
 Db 46262 AAGCGGAGTTTCTCAGAGCTGCTTCCAGGATACATCACTCAACCTCAACGACCCCTC 46203  
 Qy 1020 GGACTTTGCTGCCTAAATTCATGGACTGTACTGTGTGCGAGGAGGTGGCAAGAACATTC 1079  
 Db 46202 GGACTTTGCTGCCTAAATTCATGGACTGTACTGTGTGCGAGAGGTGGCAAGAACATTC 46143

Qy 1080 GGATTGTGGTGATGAACAATCTTTTACCAGATCGGTAAAAATGATATCAAAATATGACC 1139  
 Db 46142 GGATTGTGGTGATGAACAATCTTTTACCAGATCGGTAAAAATGATATCAAAATATGACC 46083  
 Qy 1140 TCAAGAGCTCAACCTTACAAACGGCGGCTTCCCAAGAGAGCGAGAGGCTCTTCCCA 1199  
 Db 46082 TCAAGAGCTCAACCTTACAGGCGGCGGCTTCCCAAGAGAGCGAGAGGCTCTTCCCA 46023  
 Qy 1200 CATTTAAAGACCTAGACTTCTTACAAGACATCCCTGTATGGTCTTTTGGATGCTGACA 1259  
 Db 46022 CATTTAAAGACCTAGACTTCTTACAAGACATCCCTGTATGGTCTTTTGGATGCTGAGC 45963  
 Qy 1260 TGTACAACGCTCTCTGTAAAGACCTGCAAGCTGTACTGTTGGTGTGAGAGCTTCAAGA 1319  
 Db 45962 TGTACAACGCTCTCTGTAAAGACCTGCAAGCTGTACTGTTGGTGTGAGAGCTTCAAGA 45903  
 Qy 1320 TAATGGATTACAGCTCTTGTATGTCATTAATATAGATCATGCACAAACGAGAGCCCT 1379  
 Db 45902 TAATGGATTATAGCTCTTGTATGTCATTAATATAGATCATGCACAAACGAGAGCCCT 45843  
 Qy 1380 TAAGCAGTGAACACAGTACTACTGATTGATCTCGAAGACCGGCCCCCAAAAGGCTCTGT 1439  
 Db 45842 TAAGCAGTGAACACAACTACTGATTGATCTCGAAGACCGGCCCCCAAAAGGCTCTGT 45783  
 Qy 1440 ATTCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGGTACCATGGAGACTG 1499  
 Db 45782 ATTCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGGTACCATGGAGACTG 45723  
 Qy 1500 ATGACCATATGGTGGCATCCCTGCCGGAATAGTAAAGGGGAAAGGCTCTGCTTTATA 1559  
 Db 45722 ATGACCATATGGTGGCATCCCTGCCGGAATAGTAAAGGGGAAAGGCTCTGCTTTATA 45663  
 Qy 1560 TTGGCATATTGACATTTACAGTCTTACAGGTTTGTAAAGAGTTGGAGCACTCTTGA 1619  
 Db 45662 TTGGCATATTGACATTTACAGTCTTACAGGTTTGTAAAGAGTTGGAGCACTCTTGA 45603  
 Qy 1620 AAGCCCTGGTACATGACGAGACTGTCTCAGTGCATCGCCGAGGCTTCTACGCTGAAC 1679  
 Db 45602 AAGCCCTGATACATGACGAGACTGTCTCAGTGCATCGCCGAGGCTTCTATGCTGAAT 45543  
 Qy 1680 GGTTCACGCGCTTCATGTGCAACACAGTATTAAAGAGTTCCCTTGAAGCTTCTCCTT 1739  
 Db 45542 GGTTCACGCGCTTCATGTGCAACACAGTATTAAAGAGTTCCCTTGAAGCTTCTCCTT 45483  
 Qy 1740 CCAAAAAGTTTCGGTCTGGCTCATCTTCTCGCGAGCAGGCTCCAGTGGCAACTCT 1799  
 Db 45482 CCAAAAAGTTTCGGTCTGGCTCATCTTCTCGCGAGCAGGCTCCAGTGGCAACTCT 45423  
 Qy 1800 GCATTACTTACAGCCATCGGTCTCTGGGAAACACAGGACAGTGAACAAAGGCGAG 1859  
 Db 45422 GCATTACTTACAGCCATCGGTCTCTGGGAAACACAGGACAGTGAACAAAGGCGAG 45363  
 Qy 1860 AAGTGGAGCCAGGCTTCCACTTGGTCTGCTGTATGTTTACCTCAGACTCCACCTTTGG 1919  
 Db 45362 AAGTGGAGCCAGGCTTCCACTTGGTCTGCTGTATGTTTACCTCAGACTCCACCTTTGG 45303  
 Qy 1920 AGGAAATCAGTGAAGGCTCGCTTATTCCTGACCCAGTTTCTACCTCTAGTTGGAGAGA 1979  
 Db 45302 AGGAAATCAGTGAAGGCTCGCTTATTCCTGACCCAGTTTCTACCTCTAGTTGGAGAGA 45243  
 Qy 1980 CTTTGCATAATGCTACTACAAGT 2002  
 Db 45242 CTTTGCATAATGCTACTACAAGT 45220

RESULT 11  
 AL592111  
 LOCUS Homo sapiens chromosome 1 clone RP11-124I9, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 8 unordered pieces.  
 ACCESSION AL592111  
 VERSION AL592111.3 GI:14575471



Db 92485 GATGAATTTCTTAAGGCTGAAGGAATGAAGAGAGTGGACATGGGTAATCTTTTATCCCT 92544  
QY 2824 TTTGTTAAAACAGAGGACGCATGGCTGGGAGATCATAGCCCTTCTAGGCGAATCC 2893  
Db 92545 TTTGTTAAAACAGAGGACGCATGGCTGGGAGATCATAGCCCTTCTAGGCGAATCC 92604  
QY 2884 TGTTCACCTGCGAGGCTAGTAATTAATTAATTTTCAATTTTCAATATATCTGCTG 2943  
Db 92605 TGTTCACCTGCGAGGCTAGTAATTAATTAATTTTCAATTTTCAATATATCTGCTG 92664  
QY 2944 TTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCAGAGGAGATTTT 3003  
Db 92665 TTTTCTAAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCAGAGGAGATTTT 92724  
QY 3004 TTTGCTCTTCTCATCTTTTCCACAGTGTCTCTGTTTGTGGAGCTAAAGAGGG 3063  
Db 92725 TTTGCTCTTCTCATCTTTTCCACAGTGTCTCTGTTTGTGGAGCTAAAGAGGG 92784  
QY 3064 GACACTTCTGCTCTTTAAACAGACAGTCCATATCTGTGAGGCCAGCAATATTTTCTTAA 3123  
Db 92785 GACACTTCTGCTCTTTAAACAGACAGTCCATATCTGTGAGGCCAGCAATATTTTCTTAA 92844  
QY 3124 ACTCATGGGAGACAGAGATTTCTGCTTGGTGAGTCAATTTGCTGCTCATATGCTCTTA 3183  
Db 92845 ACTCATGGGAGACAGAGATTTCTGCTTGGTGAGTCAATTTGCTGCTCATATGCTCTTA 92904  
QY 3184 CCCCCCTGCTTCATGACGAGAGTGGAAATGGGGGCTACATATGCTCTTCTCCCG 3243  
Db 92905 CCCCCCTGCTTCATGACGAGAGTGGAAATGGGGGCTACATATGCTCTTCTCCCG 92964  
QY 3244 TCTACAAGAGTGTGGTTTCCATCTCATCTCCACTCTGTGTCAGGGAGAGGGGGC 3303  
Db 92965 TCTACAAGAGTGTGGTTTCCATCTCATCTCCACTCTGTGTCAGGGAGAGGGGGC 93024  
QY 3304 CTGCTATCTCAGGACAGATTTGTAATTCCTGTTCTATCCCTTCTCTATCCACCCCTGCC 3363  
Db 93025 CTGCTATCTCAGGACAGATTTGTAATTCCTGTTCTATCCCTTCTCTATCCACCCCTGCC 93084  
QY 3364 TGATAATATGTAGCCCATACCCCAATTAATGCTATATATAGACACCCCGACGAGTTT 3423  
Db 93085 TGATAATATGTAGCCCATACCCCAATTAATGCTATATATAGACACCCCGACGAGTTT 93144  
QY 3424 CTGCTGCTCTGCTTCTGCTGCTATTTTACAGAGGAGAAATTTCTGCTATTTT 3483  
Db 93145 CTGCTGCTCTGCTTCTGCTGCTATTTTACAGAGGAGAAATTTCTGCTATTTT 93204  
QY 3484 TTTTCAATAATTTACTATTTATGATGATTTTAAAGTGTTTTATTAAGGACAGATTTCTGTTA 3543  
Db 93205 TTTTCAATAATTTACTATTTATGATGATTTTAAAGTGTTTTATTAAGGACAGATTTCTGTTA 93264  
QY 3544 GGGGTGGGAGGAATATTTGAGGAGGGCTGGGCTTTAGGAAAGGAATGGGAAAGCAAC 3603  
Db 93265 GGGGTGGGAGGAATATTTGAGGAGGGCTGGGCTTTAGGAAAGGAATGGGAAAGCAAC 93324  
QY 3604 ATTTTATTAAGTCTTACTATTTGCTTCTACTTTGTTATTTTCAAGAAATGGCAATACAA 3663  
Db 93325 ATTTTATTAAGTCTTACTATTTGCTTCTACTTTGTTATTTTCAAGAAATGGCAATACAA 93384  
QY 3664 TATAAAGTGATATATGTTTAAATGTAATAAATTTTAAAGTATTTTA 3713  
Db 93385 TATAAAGTGATATATGTTTAAATGTAATAAATTTTAAAGTATTTTA 93434

RESULT 12  
AL592424  
LOCUS Human DNA sequence from clone Rpl11-68118 on chromosome 1, complete  
DEFINITION  
AL592424  
ACCESSION  
AL592424  
VERSION  
AL592424.13  
KEYWORDS  
HTG.  
SOURCE  
human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1  
Direct Submission  
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 7, 2002 this sequence version replaced gi:19699452.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chrl  
Rpl11-68118 is from the library RPl1-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6.

FEATURES  
source  
1..159148  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="Rpl11-68118"  
/clone.lib="RPl1-11.1"  
BASE COUNT 40758 a 37585 c 37589 g 43216 t  
ORIGIN

Query Match 44.6%; Score 1654.8; DB 9; Length 159148;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1667; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 2045 CATTAAGCGCAAGCCCTCAGAGACCTGGAAACAGATTCTGCCATCTCTGTATCCCCAAG 2104  
Db 141963 CAGTAAGCGCAAGCCCTCAGAGACCTGGAAACAGATTCTGCCATCTCTGTATCCCCAAG 142022  
QY 2105 ATGTCAGCCCTTGCCTCCAGCAATGCTGAATTTCTTCTACTTGGTCATCAAAAAGGAGT 2164  
Db 142023 ATGTCAGCCCTTGCCTCCAGCAATGCTGAATTTCTTCTACTTGGTCATCAAAAAGGAGT 142082  
QY 2165 GTAATAGAAGTGAGGGAGCTGCTCCTCATCTTCTTCTGAAGAAGAACCTTCTCTCT 2224  
Db 142083 GTAATAGAAGTGAGGGAGCTGCTCCTCATCTTCTTCTGAAGAAGAACCTTCTCTCT 142142  
QY 2225 TCTCTTCTCATGAATGGGCCCTTAGTGCCCTCAGAGATTGAGGACCCGACGATCCCTC 2284  
Db 142143 TCTCTTCTCATGAATGGGCCCTTAGTGCCCTCAGAGATTGAGGACCCGACGATCCCTC 142202  
QY 2285 CACTCCAGAGTGGTGGTACGAGTTTCAACTGCCCAACCTTTGCCCTCCACTATTGAA 2344  
Db 142203 CACTCCAGAGTGGTGGTACGAGTTTCAACTGCCCAACCTTTGCCCTCCACTATTGAA 142262  
QY 2345 TTTTTCAGACCCCTTCTTCTCATGCTGGAATGGGATTGCTGACCTGGCAGCTTCT 2404  
Db 142263 TTTTTCAGACCCCTTCTTCTCATGCTGGAATGGGATTGCTGACCTGGCAGCTTCT 142322

QY 2405 TTCCCTCGCTTTGACTAGGAACCGGACTCTTAATTTCTCTCAGGACAGACTAGCTGGCA 2464  
Db 142323 TTCCCTCGCTTTGACTAGGAACCGGACTCTTAATTTCTCTCAGGACAGACTAGCTGGCA 142382  
QY 2465 CATTATCCCTACCTTAGTCTCTCTGACTCCCTGGGAAGTAATCTCTCTGTAATCTCTG 2524  
Db 142383 CATTATCCCTACCTTAGTCTCTCTGACTCCCTGGGAAGTAATCTCTCTGTAATCTCTG 142442  
QY 2525 TAAAGGTTTTGGGGGATAGGCTGTTTAAACCACCTCCAGCTCTTCTCTTC -TTTTTTT 2583  
Db 142443 TAAAGGTTTTGGGGGATAGGCTGTTTAAACCACCTCCAGCTCTTCTCTTC -TTTTTTT 142502  
QY 2584 TTTTCTGAAAAAGGAAAAAGCACACAGCACACAATTTCAAGCCATTTTCAGATCAACA 2643  
Db 142503 TTTTCTGAAAAAGGAAAAAGCACACAGCACACAATTTCAAGCCATTTTCAGATCAACA 142562  
QY 2644 CCAGAAGTGTGACAAAGTGCCTATTTCGTAGAGTTCCTCCAGAAAGCCATGGTGTAT 2703  
Db 142563 CCAGAAGTGTGACAAAGTGCCTATTTCGTAGAGTTCCTCCAGAAAGCCATGGTGTAT 142622  
QY 2704 GAAGAGAAGTAGTAGTCTCTGCTGCCAAGCAGCTCTCTTTAAACTCCCTCTCTTT 2763  
Db 142623 GAAGAGAAGTAGTAGTCTCTGCTGCCAAGCAGCTCTCTTTAAACTCCCTCTCTTT 142682  
QY 2764 GATGAATTTCTTAAAGCTGAAGGAATGAAGAGTGGGACATGGGCTAATCTTATCCCT 2823  
Db 142683 GATGAATTTCTTAAAGCTGAAGGAATGAAGAGTGGGACATGGGCTAATCTTATCCCT 142742  
QY 2824 TTTGTTAAACAGAGGAGCCATGGCTGGGAGATCATAGCCCTTCCTAGGCAGAAATCC 2883  
Db 142743 TTTGTTAAACAGAGGAGCCATGGCTGGGAGATCATAGCCCTTCCTAGGCAGAAATCC 142802  
QY 2884 TGTTCACTCCGAGCTATAGTAATATTACTATTTTGCATTTGCAATTTGAAATATATCTGGTTG 2943  
Db 142803 TGTTCACTCCGAGCTATAGTAATATTACTATTTTGCATTTGCAATTTGAAATATATCTGGTTG 142862  
QY 2944 TTTTCTTAAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCACAGAGAGATTTT 3003  
Db 142863 TTTTCTTAAATGTGAAGACTTACCAATGAATTTTAGATCATCTCCACAGAGAGATTTT 142922  
QY 3004 TTTGCTCTCTCATCTTTTCCACAGTGTTCCTGTTTGTGGAGCTAAGGTAAGAGAGG 3063  
Db 142923 TTTGCTCTCTCATCTTTTCCACAGTGTTCCTGTTTGTGGAGCTAAGGTAAGAGAGG 142982  
QY 3064 GACACTTCTGCTGTTTAAACAGACAGTCCATATCTGTGAGGCGCCAGCAATATTTCTTAA 3123  
Db 142983 GACACTTCTGCTGTTTAAACAGACAGTCCATATCTGTGAGGCGCCAGCAATATTTCTTAA 143042  
QY 3124 ACTCATGGGGAGACAGAGATTTTCCCTTGTGTGAGGTGATTCCTGTCATATGTCCTA 3183  
Db 143043 ACTCATGGGGAGACAGAGATTTTCCCTTGTGTGAGGTGATTCCTGTCATATGTCCTA 143102  
QY 3184 CCCCCCTGCTTCATGACGGGAAGTTGGAATGGGGGCTACATATGCCCTCTCTCCCG 3243  
Db 143103 CCCCCCTGCTTCATGACGGGAAGTTGGAATGGGGGCTACATATGCCCTCTCTCCCG 143162  
QY 3244 TCTACAAGAGTTGTGGTTTTCATCTGATCCCTCCACCTCTCTCAGGCGAAGAGGGGC 3303  
Db 143163 TCTACAAGAGTTGTGGTTTTCATCTGATCCCTCCACCTCTCTCAGGCGAAGAGGGGC 143222  
QY 3304 CTGGTATCTCAGGAGATTTGTAATCTCTGTTCTATCCCTCTCTATCCCAACCCCTGCT 3363  
Db 143223 CTGGTATCTCAGGAGATTTGTAATCTCTGTTCTATCCCTCTCTATCCCAACCCCTGCT 143282  
QY 3364 TGATAATATGTTAGCCCATACCCCAATAACTGTCTATATTAGACACCCCGACCCAGTTT 3423  
Db 143283 TGATAATATGTTAGCCCATACCCCAATAACTGTCTATATTAGACACCCCGACCCAGTTT 143342  
QY 3424 CTGGCTGCTGCTTCTGCTGCTATTTTACAGAGAGGAAGTAATCTGCTATTTT 3483  
Db 143343 CTGGCTGCTGCTTCTGCTGCTATTTTACAGAGAGGAAGTAATCTGCTATTTT 143402  
QY 3484 TTTTTCATAATTTACTATTATGATGATTTTAAAGTGTGTTTTTAAAGGACAGAGTCTCTGTTA 3543

Db 143403 TTTTTCATAATTTACTATTATGATGATTTAAAGTGTGTTTTTAAAGGACAGAGTCTCTGTTA 143462  
QY 3544 GGGGTGGGAGGGAATATTTGAGGAGGGCTGGTCTTAGGGAAGGAATGGGGAAGCAAC 3603  
Db 143463 GGGGTGGGAGGGAATATTTGAGGAGGGCTGGTCTTAGGGAAGGAATGGGGAAGCAAC 143522  
QY 3604 ATTTTATTAAAGTGTACTATTGCTCTACTTGTGCTTGTATTGTTTCAGAAATGGCAATACAA 3663  
Db 143523 ATTTTATTAAAGTGTACTATTGCTCTACTTGTGCTTGTATTGTTTCAGAAATGGCAATACAA 143582  
QY 3664 TATAAAGTCATATATGTTTAAATGTAATAAATCTTAAATGAGTTATTTA 3713  
Db 143583 TATAAAGTCATATATGTTTAAATGTAATAAATCTTAAATGAGTTATTTA 143632

## RESULT 13

AC009564/c

LOCUS

DEFINITION

AC009564

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

## TITLE

JOURNAL

REFERENCE

AUTHORS

AC009564 184090 bp DNA linear HTG 24-AUG-2002  
Homo sapiens chromosome 10 clone RP11-413M23 map 10, WORKING DRAFT  
SEQUENCE, 36 unordered pieces.

AC009564  
AC009564.5 GI:8072465

HTGS: HTGS\_PHASE1; HTGS\_DRAFT.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 184090)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 184090)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baker, J., Baldwin, J., Barna, N., Becker, R., Benn, J., Brown, A.,

Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,

Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,

Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,

Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,

Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,

Meldrim, J., Mollif, M., Morris, W., Morrow, J., Mychaleckyj, J.,

Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,

Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,

Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,

Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 184090)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,

Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., Lhocque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Levine, R., Liebow, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Willson, B., Wu, X., Wyman, D., Ye, W. J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (24-AUG-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7637235.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L1692  
Center clone name: 413\_M\_23  
----- Summary Statistics  
Sequencing vector: M13; M77815; 96% of reads  
Sequencing vector: Plasmid; n/a; %0.1% of reads  
4.25090720580612Chemistry: Dye-primer-amersham; 4% of reads  
Chemistry: Dye-terminator Big Dye; 96% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 151999 bases at least Q40  
Consensus quality: 164409 bases at least Q30  
Consensus quality: 171030 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 180590; sum-of-contigs  
Quality covera.

NOTE: This is a 'working draft' sequence. It currently consists of 36 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1060: contig of 1060 bp in length  
\* 1061 1160: gap of 100 bp  
\* 1161 2375: contig of 1215 bp in length  
\* 2376 2475: gap of 100 bp  
\* 2476 3802: contig of 1327 bp in length  
\* 3803 3902: gap of 100 bp  
\* 3903 5236: contig of 1334 bp in length  
\* 5237 5336: gap of 100 bp  
\* 5337 7037: contig of 1701 bp in length  
\* 7038 7137: gap of 100 bp  
\* 7138 8383: contig of 1246 bp in length  
\* 8384 8483: gap of 100 bp  
\* 8484 9908: contig of 1425 bp in length  
\* 9909 10008: gap of 100 bp  
\* 10009 11137: contig of 1129 bp in length  
\* 11138 11237: gap of 100 bp  
\* 11238 13218: contig of 1981 bp in length  
\* 13219 13318: gap of 100 bp  
\* 13319 16067: contig of 2749 bp in length  
\* 16068 16167: gap of 100 bp  
\* 16168 18833: contig of 2666 bp in length  
\* 18834 18933: gap of 100 bp  
\* 18934 22378: contig of 3445 bp in length  
\* 22379 22478: gap of 100 bp  
\* 22479 24946: contig of 2468 bp in length  
\* 24947 25046: gap of 100 bp  
\* 25047 26306: contig of 1260 bp in length  
\* 26307 26406: gap of 100 bp  
\* 26407 29679: contig of 3273 bp in length  
\* 29680 29779: gap of 100 bp  
\* 29780 33599: contig of 3820 bp in length  
\* 33600 33699: gap of 100 bp  
\* 33700 36502: contig of 2803 bp in length  
\* 36503 36602: gap of 100 bp  
\* 36603 39914: contig of 3312 bp in length  
\* 39915 40014: gap of 100 bp  
\* 40015 42987: contig of 2973 bp in length  
\* 42988 43087: gap of 100 bp

\* 43088 46873: contig of 3786 bp in length  
\* 46874 46973: gap of 100 bp  
\* 46974 50544: contig of 3571 bp in length  
\* 50545 50644: gap of 100 bp  
\* 50645 53900: contig of 3256 bp in length  
\* 53901 54000: gap of 100 bp  
\* 54001 58793: contig of 4793 bp in length  
\* 58794 58893: gap of 100 bp  
\* 58894 64521: contig of 5628 bp in length  
\* 64522 64621: gap of 100 bp  
\* 64622 69423: contig of 4802 bp in length  
\* 69424 69523: gap of 100 bp  
\* 69524 75858: contig of 6335 bp in length  
\* 75859 75958: gap of 100 bp  
\* 75959 82538: contig of 6380 bp in length  
\* 82539 82638: gap of 100 bp  
\* 82639 89982: contig of 7344 bp in length  
\* 89983 90082: gap of 100 bp  
\* 90083 96628: contig of 6546 bp in length  
\* 96629 96728: gap of 100 bp  
\* 96729 106892: contig of 10164 bp in length  
\* 106893 106992: gap of 100 bp  
\* 106993 118329: contig of 11337 bp in length  
\* 118330 118429: gap of 100 bp  
\* 118430 130262: contig of 11833 bp in length  
\* 130263 130362: gap of 100 bp  
\* 130363 141462: contig of 11100 bp in length  
\* 141463 141562: gap of 100 bp  
\* 141563 154056: contig of 12494 bp in length  
\* 154057 154156: gap of 100 bp  
\* 154157 164998: contig of 10842 bp in length  
\* 164999 165098: gap of 100 bp  
\* 165099 184090: contig of 18992 bp in length.  
FEATURES  
Location/Qualifiers  
1. 184090  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10"  
/clone="RP11-413M23"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 1060  
/note="assembly\_fragment"  
misc\_feature 1161..2375  
misc\_feature /note="assembly\_fragment"  
2476..3802  
misc\_feature /note="assembly\_fragment"  
3903..5236  
misc\_feature /note="assembly\_fragment"  
5337..7037  
misc\_feature /note="assembly\_fragment"  
7138..8383  
misc\_feature /note="assembly\_fragment"  
8484..9908  
misc\_feature /note="assembly\_fragment"  
10009..11137  
misc\_feature /note="assembly\_fragment"  
11238..13218  
misc\_feature /note="assembly\_fragment"  
13319..16067  
misc\_feature /note="assembly\_fragment"  
16168..18833  
misc\_feature /note="assembly\_fragment"  
18934..22378  
misc\_feature /note="assembly\_fragment"  
22479..24946  
misc\_feature /note="assembly\_fragment"  
25047..26306  
misc\_feature /note="assembly\_fragment"  
clone\_end:77  
vector\_side:left  
26407..29679

Query Match	43.6%;	Score 1618.2;	DB 2;	Length 184090;
Best local Similarity	92.9%;	Pred. No. 0;		
Matches 1769;	Conservative 0;	Mismatches 113;	Indels 23;	Gaps 6;

  

Qy	105	CCACAGCAGCAGCTTAAAGCTTACTCTCTGTAAGAGGGGAAAGTATCCCTCTGTGGAAAGC	164
Db	118262	CCACAGAGCAAGTTTAAAGTTTTTTTTTTTGGAAAGGGGAAAGTTCCCTCTGGGAAAC	118203
Qy	165	GGTTAAACTTGTGGAGGGGGTGCAGGAGCTGAGTGTCTTCCCC--ATGCCAGCGCAATGGT	222
Db	118202	AGTTAAATTTGTGAGGGGGTGCAGAGCTAAGTTCCTCCCCCATGCCAGGGGAATGGT	118143
Qy	223	GTGGCCTTGAGTGTCTCAGGAGCGGGCTGCACGTGTCTGAGGAGGCCGAGGGGGCG	282
Db	118142	GTGGATTTAGTGTGTCCAGCAGCGGGTTTCAATGTTTGTAGGGAAGCCCTGGAGGGGGG	118083
Qy	283	GGAGGTCGGCCACACA--GAACGGGGTTCGTAAGAGAGCTTGTGGAAGATTCGATTCC	339
Db	118082	GGGAGGGCCCTCCACAGAAAGTGGTTTTTATAAAAGACATTTGGGAAGATTCGATTCA	118023
Qy	340	GAGAAGAG--AAGAACCAGTTGAAGAGAG--CCAGGCGCTGAGGGGAGGGGGCTGCT	397
Db	118022	GAGAAGGAAGAGCGGGATTAAGAGAGCCACAGGCCCTTGAGGGGAGGGGGCTGCC	117963
Qy	398	AAGATGGCGTCGGCTCTCTCCGGGCGCTGCTGCTCGGTTCGGTTTTTCATCCTTTGATCCC	457
Db	117962	AAGATGGCGTCGGCTCTCTCCGGGCGCTGCTGCTCGGTTCGGTTTTTC--TCGTTGATCCC	117904
Qy	458	CGGTCCTCTCTGTAACCTTGTCTTCAGCATCTCGAATCAAGAGACCCATGGCATTCTGAG	517
Db	117903	CGGTCCTCTCTGTAACCTGCTCTCAGCATCTCGAATCAAGAGTCCCATGGCATCTGAG	117844
Qy	518	GTGCTTATGCTCTGGCATGCCCATCAAGAAAATAGCCATAGAAGTGTGTGATTCCTCA	577
Db	117843	GTGCTTATGCTCTGGCATGCCCATCAAGAAAATAGCCATAGAAGTGTGTGATTCCTCA	117784
Qy	578	CGAGAGACAATATAAAGACAACTCATCAGCCTTGAAGGTGCCATCCAGTTAGGC	637
Db	117783	GGAG-----GGACAACTCATCAGCCTTGAAGGTGCCATCCAGTTAGGC	117739
Qy	638	ATTACCCACACTGGGGAGCCTGAGTACCAACACAGAGCGTGTATGTCCTCATGCAAGAT	697
Db	117738	ATTACCCACACTGGGGAGCCTGAGTACCAACACAGAGAGTGTATGTCCTCATGCAAGAT	117679
Qy	698	TTCTACGTGTTGAGAGTATCTTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCAT	757
Db	117678	TTCCACATGTTGGAGAGTATCTTTTCCAGTGAAGGAGCAACCTGACCCCTGCTCAT	117619
Qy	758	CACTAACATGACTTTCGTTTCAAGACCTATGCACCTGTTCGCTTCGCTACTTCCGGAG	817
Db	117618	CACTAACATGCTTTCGTTTCAAGACCTATGCACCTATGCACCTATGCCTACTTCTGGAG	117559
Qy	818	CTATTTGATTCGGCCCGGATGATTCTGTAATTCCTCTGCAAGTGTGAGCGGCTGATTGA	877
Db	117558	CTATTTGATTCGGCCCGGATGATTCTGTAATTCCTCTGCAAGTGTGAGCGGCTGATTGA	117499
Qy	878	CTCTGTAGCTCTGAGCTAGTGTTCCTATCTATGTGTCCAGCAGCATGATGATTCATT	937
Db	117498	CTCTGTAGCTCTGAGCTAGTGTTCCTATCTATGTGTCCAGCAGCATGATGATTCATT	117439
Qy	938	ATTAAAGACAGTCCAACATAAAGAGGGGAATTTCTGCAGAGCTGCTTCCAGGATACTAC	997
Db	117438	GTTAAGACAGTCCGACATAAAGAGGGGAGTTTCTGCAGAGCTGCTTCCAGGATACTAC	117379
Qy	998	ATGAACCTCAACAGACCCCTCGGACTTTGCTGCCTAAATTCATGGACTGTACTGTGG	1057
Db	117378	ATAACCTCAACAGACCCCTCGGACTTTGCTGCCTAAATTCATGGACTGTACTGTGG	117319
Qy	1058	CAGCAGGTGGCAAGAACATTCGGATTGTGGTGATGAACAATCTTTTACCAAGATCGGTA	1117
Db	117318	CAGCAGGTGGCAAGAACATTCGGATTGTGGTGATGAACAATCTTTTACCAAGATCGGTA	117259
Qy	1118	AAATGCATATCAAAATATGACCTCAAGGCTCAACCTACAAAGCGGGCTTTCCAGAAA	1177

[illegible]

BC031774 2553 bp mRNA linear ROD 07-AUG-2002  
Mus musculus, phosphatidylinositol-4-phosphate 5-kinase, type 1  
beta, clone MGC:25310 IMAGE:4503697, mRNA, complete cds.

BC031774.1 GI:21619322

MGC.  
house mouse.

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2553)

# AUTHORS TITLE JOURNAL

Strausberg, R.  
Direct Submission  
Submitted (06-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

## REMARK COMMENT

NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcaps@remail.nih.gov](mailto:gcaps@remail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
Yoon, V.S., Kowls, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 31 Row: 1 Column: 14  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 5679330.

## FEATURES source

Location/Qualifiers  
1..2553  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="MGC:25310 IMAGE:4503697"  
/tissue\_type="Eye, retina, mouse strain C57BL/6"  
/clone\_lib="NIH\_MGC\_94"  
/lab\_host="DH10B"  
/notes="Vector: pCMV-SPORT6"  
408..2048  
/codon\_start=1  
/product="phosphatidylinositol-4-phosphate 5-kinase, type  
1 beta"  
/protein\_id="AAH31774.1"  
/db\_xref="GI:21619323"  
/db\_xref="LocusID:18720"

## CDS

/translations="MASASSGPAAGFSSLDAGAPAGTAASGKIKRATVSEGPSASVM  
PVKIGHRSDVSGEYTKTSSALKAQLQITHTVGLSLRKPDRVLMQDFYVE  
SIFPSEGSLTPAHNIDFRFTYAPVAFRFLGIRPDYLYSLCEPLIELSN  
SGAGSLFYVSDDEFLITVQKAEFLQKLPGYMNLNPNRLLPFYGLYCVQ  
AGKNTIRIVYMNLLPRSVKMHKDYLDKSTYKRASQKEREKTLPTFKDPLQIDP  
DGLFDADMTSALCKLORDCLVQSFIMDYSLIMSINHMDHAOREPTSDTOYSAD  
TRPAQKALYSTAMSIQGEARRGVETEDHMGIPARNKGERLLLYIGLIDILQ  
SYRFVKLEHSKALVHDGTVSVHRPGFYAERFQRMCMNTYFKKIPKSPFKFRS  
GPSFRSRSGPSGNSCTSQLMASGHEHRAVTTKAEVDPVHLGRDPDLPQTPPLEEISE  
GSPVGPSPVVGQPLQILNLSTLEKLDVAESEFTH"

## BASE COUNT ORIGIN

Query Match 41.6%; Score 1543.4; DB 10; Length 2553;  
Best Local Similarity 78.7%; Pred. No. 0;  
Matches 2050; Conservative 0; Mismatches 471; Indels 84; Gaps 14;

Qy 1 ATTAACAGCGCGTGTAGGAAGCAGGAGAGGGCGTTCGCTCCCTTTGGG---ACTTT 57  
Db 30 ATAAACAGGCGATGGCTGGGAGGTGAGGAGGGGGGCGATTCGCTTTGGGGCACCCTC 89  
Qy 58 TCATCGCTCGTTTTTTTTCAGATGTGGCTTGGCTGGCGCAAGGTCGCCAGCAGCAGC 117  
Db 90 AGGTGCGCTAGTTCCTTTAGTTTGTCTCTCCGAGACGCGGACTCCCGGGGCGCGG 149  
Qy 118 TTAGCTTACTCTCTCTGTGAAGGGGAAGATATCCCTGTGGAAAGCGGTTAAACTGTG 177  
Db 150 CTCA-AGTGTTCCTCTGTGTGAAGAGAGGTAGCTTTGGCGGAGAGCAGTTCAAGTGTCT 208  
Qy 178 GAGGGGTGGGAGCGTGTCTTCCCATGCCAGCGAATGTGTGGCTTCAGCTGG 237

Db 209 AGAAGGTGCGAGACTGCAGCTCTCCCATGCGGGCAAGTGGTGGCGGTCTCTGAGCTTG 268  
Qy 238 TCCAGGAGCGCGCTCGACGTCTCTGAGGAGAGCCCGGAGGGGGGAGGTGGCCACCA 297  
Db 269 TCTTGGAGCTTGAACAACGCGCTGTGA-----GAAGCCCGCC 304  
Qy 298 GAACGCGGTTCTGTAAAGAGACGTTGGGAAGATTCGATTCCGAGAGAAGAGAACCGG 357  
Db 305 GAACGCGGTTCTGTGAAGAGACGTTGGGAAGATTCGATTCCGAGAGAAGAGAACCGG 364  
Qy 358 ATTGAAGAGAGACGAGCGCTGAGGGGAGGGGCTGCTAAGATGGGTGGCTCTCTC 417  
Db 365 ATTGAAGAGAGACGAGCGCTGAGGGGAGGGGCTGCCAAGATGGGTGGCTCTCTC 424  
Qy 418 CGGCGCGTCTGCTTCGGTTCGCTTTTTCATCTCTTGTATCCCGCGGCTCTCTGTACCTT 477  
Db 425 AGGCGCAGCG---CGGCGCGGTTTTCATCTCTTGTATCCCGCGGCTCTCTGTACCTT 479  
Qy 478 GTCCTCAGCATCTGGAATCAAGAGACCATGGCATCTCAGGTGCTTATGCTCTGGCAT 537  
Db 480 -GCAGCAGCATCTGGAATCAAGAGAGACCATGGCATCTCAGGTGCTTATGCTCTGGCAT 538  
Qy 538 GCCCATCAAGAAATAGGCCATAGAAGTGTGATTTCCTCAGGAGACACAATATAAAAA 597  
Db 539 GCCTGTTAAAAAATAGGCCATCGAAGTGTGATTTCCTCTGGAGAGACTACATACAGAA 598  
Qy 598 GACAACTCATCAGCCTTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAG 657  
Db 599 GACAACTCATCAGCCTTTGAAGGTGCCATCCAGTTAGGCATTTACCCACACTGTGGGAG 658  
Qy 658 CTGAGTACCAAAACAGAGCGTGTCTCATGCAAGATTTCTACGTTGGTGGAGATAT 717  
Db 659 CTGAGTACCAAAACAGAGCGTGTCTCATGCAAGATTTCTACGTTGGTGGAGATAT 718  
Qy 718 CTCTCTTCCAGTGAAGGGAGCAACTGACCCCTCTCATCTACATCAATGACTTTCTGTTT 777  
Db 719 CTCTCTTCCAGTGAAGGGAGCAACTGACCCCTCTCATCTACATCAATGACTTTCTGTTT 778  
Qy 778 CAAGACCTATGCACCTGTGCTTCCGCTACTTCCGGAGCTATTGGTATCCGCCCGCA 837  
Db 779 CAAGACCTATGCACCTGTGCTTCCGCTACTTCCAGAGAGCTCTTTGGCATCCGCCCTGA 838  
Qy 838 TGATTACTGTATCCCTCTCGAGTGGCGCTGATTGAACCTCTAGCTCTGGAGCTAG 897  
Db 839 TGATTACTGTATCCCTCTCGAGTGGCGCTGATTGAACCTCTAGCTCTGGAGCTAG 898  
Qy 898 TGGTTCCCTATCTATGTGTCCAGGACGATGAGTTCAATTATTAAGACAGTCCCAACATAA 957  
Db 899 TGGTTCCCTATCTATGTGTCCAGTGTGATGAATTCATCATTAAGACCGTCCAGCATAA 958  
Qy 958 AGAGCGGAATTTCTGCAGAGAGCTGCTTCCAGGATFACATGAACCTCAACAGAACCC 1017  
Db 959 AGAAGCAGAAATTTCTGCAGAAAGTTGTTCCAGGATFACATGAATCTTAACCAAAACC 1018  
Qy 1018 TCGGACTTGTGCTGCTTAATTTCTATGAGCTGTACTGTGTGAGGAGCTGGCAAGAAAT 1077  
Db 1019 TCGGACTTGTGCTGCTTAATTTCTATGAGCTGTACTGTGTGCAAGCAGCGCGCAAGAAAT 1078  
Qy 1078 TCGGACTTGTGCTGATGAACAATCTTTCCTCGGTGCTGCTCAAAATGATATGAATATGA 1137  
Db 1079 ACAGATGTGTGATGAACAATCTTTCCTCGGTGCTGCTCAAAATGATATGAATATGA 1138  
Qy 1138 CCTCAAGGCTCAACCTACAAACCGCGGCTTCCAGAAAGAGCGAGAGACCTCTTCCC 1197  
Db 1139 CCTGAAGGTTCAACTTACAAGCGCGGCTTCTCAGAAAGAAACGAGAGAAACCTCTCCC 1198  
Qy 1198 CACATTTAAAGACCTAGACTTCTTACAAGACATCCCTGATGGTCTTTTGGATGCTGGA 1257  
Db 1199 CACCTTTAAAGACCTGGACTTCTTACAAGATATCCCTGACGCGCTATTTTGGATGCTGA 1258  
Qy 1258 CATGTACACGCTCTCTCTGAAGACCTCGACGCTGACTTGTGGTGGTGGTGGTGGTGGT 1317  
Db 1259 CATGTACAGTGTCTGTGTGAAGACTCTTACAGCGTGTGTGGTGGTGGTGGTGGTGGT 1318







Db 59 ATAAACAGCAGTGGCTGGGAGCGTGAAGGAGGGGGGCAATTCGCTTTTGGGGCACGTC 118  
Qy 58 TCATGCGCTCGTTTTTTTTCAGATGTGGCTGGTCTGGGCGCAAGTCCCGAGCAGCCAGC 117  
Db 119 AGGTGGCTAGGTCTTCTTTAGGTTTGTCTATCCGAGACGCGGACTCCCGGGGCCCGG 178  
Qy 118 TTAAGCTTACTCTCTGTGAAAGGGGAAAGTATCCCTGTGGAAGCGGTTAAACTTGTG 177  
Db 179 CTCA-AGTGTCTCTGTGTAGAGAGGTAGCTTTGGCGGAGAGCAGTTCAAGTTGCT 237  
Qy 178 GAGGGGTGGGAGCGTGAAGTTCTCCCATGCCAGCGAATGTGTGGCTTGAGCTGG 237  
Db 238 AGGAAGGTGGCAGAGCTGCGCTCCCGCATGCCGAGCGGCAAGTGTGGCTCGCTGAGCTG 297  
Qy 238 TCCAGGAGCGGCTCGAGCTGTCTGAGGAGGCGCGGAGGGGGCGGAGGTGCGCCACA 297  
Db 298 TCCTGGAGCCTGAACACGCGTGTGA-----GAGCGCGCC 333  
Qy 298 GAACCGGGTTCTCTAAAGACAGCTTGGGAAGTTCGATCCGAGAAGAGGAAGACCGG 357  
Db 334 GAACCGGGTTCTGTGAAGACAGCTGGGAAGATTCTGATCCGAGAAAGGAAGCGG 393  
Qy 358 ATTGAAGAGAGCAGCGCGCTGAGGGGAGGGGCTGCTAAGATGGCTGGCGCTCCTC 417  
Db 394 ATTCAGAGGAGCGAAGCGCTGAGGGGAGGGGCTGCCAAGATGCGCTCCGCTCCTC 453  
Qy 418 CGGCGCTGCTCTTCGCTGGTTTTTTCATCCTTTGATCCCGCGTCCCTCTCTGACCTT 477  
Db 454 AGGCGCAGCG---CGGCGCGGGTTTTTCATCCCTTGTATGCCGGGCGCTGCTGTGATCC-- 508  
Qy 478 GTCCCTCAGCATCTGGAATCAAGACACCATGGCATCTGAGTGCCTTATGCTCTGCGAT 537  
Db 509 -GCAGCAGCATCTGGATCAAGACAGCCACAGTATCTGAGGGACCTTCCGGCTCTGTAT 567  
Qy 538 GCCCATCAAGAAAATAGGCCATAGAAGTTGTATTCCTCAGGAGAGACACATATAAAA 597  
Db 568 GCCTGTATAAAAAATAGCCATCGAAGTGTGTATTCCTCTGGAGAGACTACATACAGAA 627  
Qy 598 GACAACTCATACGCTTGAAGGTGCCATCCAGTTAGGCACTACTCACACTGTGGGAG 657  
Db 628 GACAACTCATACGCTTGAAGGTGCCATCCAGTTAGGCACTACTCACACTGTGGGAG 687  
Qy 658 CCTCAGTACCAAAACAGAGCGTGTATGCTTCATCAAGATTCTACGTGGTTGAGAGTAT 717  
Db 688 CCTGAGCACCACAGCGGTGAGCTGCTGATCGAAGACTTCTACGTGGTGGAGAGCAT 747  
Qy 718 CTTCTTTCCAGTGAAGGGAGCAACCTGACCCCTGCTCATCAATGACTTTTCGTTT 777  
Db 748 CTTTTTCCCAAGTGAAGGCAGCAACCTGACACCGGCTCACCATTTACAATGACTTTTCGATT 807  
Qy 778 CAAGACCTATGCACCTGTGGCTTCGCTACTTCGGGAGGTATTTGGTATCCGGCCGGA 837  
Db 808 CAAGACCTATGCAGCTGTGGCTTCGCTTCTTACAGAGAGCTTTTGGCATCCGGCCCTGA 867  
Qy 838 TGATTACTTGTATTCCTCTCAGTGAAGCGCTGATTGAATCTGTAGCTCTGGAGCTAG 897  
Db 868 TGATTACTTGTACTCCCTTTGCACTGAGCCATGATTGAATCTTCCAAATCTGGAGCTAG 927  
Qy 898 TGGTTCCCTATTTATGTGTCCAGCAGCATGAGTTTCATTATTAAAGACAGTCCCAACATAA 957  
Db 928 TGGTTCCCTCTTCTATGTGTCCACTGATGAATTCATCATTTAAGACCGTCCAGCATAA 987  
Qy 958 AGAGCGGGAATTTCTGCAGAGCTGCTTCAGGATATCTACATGAACCTCAACAGAGACCC 1017  
Db 988 AGAAGCAGAAATTTCTGCAGAAAGTTGCTTCCAGGATCTACATGAATCTTACCAAAAACCC 1047  
Qy 1018 TCGGACTTTGTGCTGCTAAATTTCTATGGACTGTACTGTGTCAGGAGGTGGCAAGACAT 1077  
Db 1048 TCGTACTTTGCTGCCCAATTTTAAGATTGATGTTGTCAGACAGCGCGGCAAGACAT 1107  
Qy 1078 TCGGATTGTGGTATGAACATCTTTTACCAGATCGGTAAAAATGCATATCAAAATGA 1137  
Db 1108 ACGAATTTGTGTGATGAACAATCTCTTGGCTCGGTCAAGTCAAAATGCATATGAATATGA 1167

Qy 1138 CCTCAAAGCGCTCAACCTACAAACGGCGGGCTTCCAGAAAGAGCGAGAGAGCCCTCTCTCC 1197  
Db 1168 CCTGAAGGGTTCAACTTACAAAGCGGGGCTTCTCAGAAAGAACGAGAGAAACTCTCCC 1227  
Qy 1198 CACATTTAAAGACCTAGACTTTTACAAGACATCCTCATGTCTTTTGTGGATGCTGA 1257  
Db 1228 CACCTTTAAAGACCTGGACTTCTTACAAGATATCCTCAGCGGCTATTTTGTGGATGCTGA 1287  
Qy 1258 CATGTACACGCTCTCTGTAAAGACCTCGACGCTGACTGTTTGTGGTGCAGAGCTTCAA 1317  
Db 1288 CATGTACAGTCTCTGTGAAGACTCTACAGCTGACTGTTTGTGGTGCAGAGCTTCAA 1347  
Qy 1318 GATTAATGGATTACAGCTCTTTGATGTCAATCCATAATATAGATCATGTACACAGAGACC 1377  
Db 1348 GATAATGGACTACAGCTCTTTGATGTCCATCCACAATAGGACCATGACAGAGGAGCC 1407  
Qy 1378 CTTAAGCAGTCAAAACACAGTACTCAGTTGATCTGAAGACGGGCCCCCAAAAGGCTCT 1437  
Db 1408 CACGAGCAATGACACAGTACTCGGCTGACCGCGCAGACAGCCCCACAGAGGCGCT 1467  
Qy 1438 GTATTCCACAGCCATGGAATCCATCCAGGAGAGGCTCGAGGGGTGTACCATGGAGAC 1497  
Db 1468 CTATTCCACAGCTATGGAATCCATCCAGGCGAGGCTCGACGGGCGGCACTGTGGAGAC 1527  
Qy 1498 TGATGACCATATGGGTGGCATCCCTGCCCGGAATAGTAAAGGGGAAAGGCTTCTGCTTTA 1557  
Db 1528 TGAGGACCATATGGGTGGCATCCCTGCCCGGAATAACAAAGGAGAAAGGCTCTGCTTTA 1587  
Qy 1558 TATTGGCATATGACATTTACAGTCTTACAGTCTTACAGTCTTCTAAGAGTTTGGACACTCTTG 1617  
Db 1588 TATTGGCATATGATATTTCTTCACTTCACTGCTTCAAGTCTTCTAAGAGTTTGGACACTCTTG 1647  
Qy 1618 GAAAGCCCTGTATACATGAGGAGACACTGTCTCAGTGCATCGCCAGGCTTCTACGCTGA 1677  
Db 1648 GAAAGCACTGTCTCATGATGGGATACGCTGTCACTGATCGTCCAGGCTTCTATGCTGA 1707  
Qy 1678 ACGGTTCCAGCGCTTTCATGTGCAACACAGTATTTAAGAGATTTCCCTTGAAGCCCTTCC 1737  
Db 1708 CGGTTTCCAGCGCTTTCATGTGCAACACAGTGTTCAGAGATTTCCCTTGAAGCCCTTCC 1767  
Qy 1738 TTCCAAAAAGTTTCGGTCTGCTCATCTTCTCTCGGGAGCAGGCTCCAGTGGCAACTC 1797  
Db 1768 TACCAAAAAGTTTCGGTCTGCGCTGCTTCTCTCGGGATCAGGCCCGGCAACTC 1827  
Qy 1798 CTGATTACTTTACCAGCATCGCTCTGTGGGAAACACAGGACACAAAGTGAACAAGGC 1857  
Db 1828 CTGC---ACCTCCAGCTGATGGCTCTGTGGGAAACACAGAGACACAAGTGACCAAGGC 1884  
Qy 1858 AGAAGTGGAGCGCGCTTCACTTGTGCTGCTGATGTTTACCTCAGATCCACCTTT 1917  
Db 1885 GGAAGTGGAGCGCAGATGTACACCTTGGGCGTCTGATGCTTACCTCAGATCCACCTTT 1944  
Qy 1918 GGAGAAATCAGTGAAGGCTCGCTTATCTTACCGCCAGTTTCTCACCTTAGTTGGAGA 1977  
Db 1945 GGAGAAATAGTGAAGGTTTACCTTGTTCGCCCCAGTTTCTCACCTTAGTTGGAGA 2004  
Qy 1978 GACTTTGCAATGCTTAACATAAGTACAACTTGGAAAGCTTGAAGTTGAGAGTCTAGA 2037  
Db 2005 ACCTTTGCAATACTAAATTTGAGTTCAACCTTGGAAAGCTTGATGTTGAGAGTCTAGA 2064  
Qy 2038 GTTCACCCATTAAGCGAAAGCCTCAGAAAGCTTGGAAACAGATTTCTGCCATCTCTGGA 2097  
Db 2065 GTTCACCCACTGAGCAGTGTAGCCTCAGAAAGCTTGGAAAGCTTGTGAGAGTCTAGA 2124  
Qy 2098 TCCCAAGATGTACGCTTGGCCAGCAATGCTGAATTTTCTTCTTACCTTGTGCTCAAAA 2157  
Db 2125 TCCCAAGATGTACGCTTGGCCAGCAATGCTGAA---TCGCTTCTTGTGCTCAACA 2181  
Qy 2158 AAGGATGTAAATAGAGTGAAGGAGCTGCTCCTCCTCCTTCTTCTTGAAGAGAACCTT 2217  
Db 2182 GAGGAGAAATGTGAGCTACGAGGCTCTGAC-----ACGT 2219

